WPI-CS-TR-09-09 October 2009

A Multi-sensorial HRI Interface for Teleoperated Robots

By

Paulo Gonçalves de Barros

Robert W. Lindeman

Matt O. Ward

Computer Science Technical Report Series

WORCESTER POLYTECHNIC INSTITUTE

Computer Science Department

100 Institute Road, Worcester, Massachusetts 01609-2280

Abstract

Robot teleoperation interfaces are mostly graphics based nowadays, that is, all the information the operator receives is presented on a screen. Complex robot interfaces may lead to operator cognitive overload. The use of other senses to receive part of the data sensed and transmitted by the robot may help reduce this overload and thus enhance the performance of the operator. This paper aims at measuring the benefits of using vibro-tactile feedback in Human-Robot Interaction (HRI) interface design specifically for an urban search and rescue (USAR) system. Our hypothesis is that the use of collision-proximity feedback interfaces (CPFs) should lead to an improvement of an HRI system performance by increasing the operator's situation awareness (SA) and reducing cognitive load. Additionally, it should be flexible enough to be adapted to different HRI USAR systems.

A user study encompassing a search task was performed to evaluate this new interface. An in-between subjects experiment tested the effect of both graphical and vibro-tactile CPF interfaces on performance in a simple search task in a virtual collapsed-building environment. Performance and situation awareness were measured based on task time, number of collisions with the environment, number of objects found and correct report of environment using sketchmaps.

First, the results of this research highlight the importance of a homogeneity verification in the experiment groups, which is generally not reported in most research results but that can drastically affect the results of an experiment.

Second, our results indicate how previous experience can affect subjects' performance. Videogame experience seemed to have a slight impact on the performance of subjects.

Third and most importantly, our results have indicated that the use of both vibro-tactile and graphical feedback interfaces may improve operator's performance in a search environment, and may indicate an increase in operator situation awareness (SA). Future enhancements in our system to better approximate it to a real robot control experience will help us consolidate the initial results obtained here.

Table of Contents

List of Figures

List of Tables

1. Introduction

Human-robot interaction (HRI) commonly happens through the use of a graphical interface and ordinary input devices such as a joystick or gamepad, mouse and keyboard. By using the computer screen to translate the data sensed by the robot into information, the operator remotely performs tasks as if occupying the same space as the robot. Commands performed by the operator through the input controls are transmitted to the robot and converted into actions. The actions change the state of the robot and potentially the state of the physical environment, and therefore the data being sensed and monitored by the system. This interaction cycle, as shown in Figure 1, continues indefinitely as a mission is carried out.

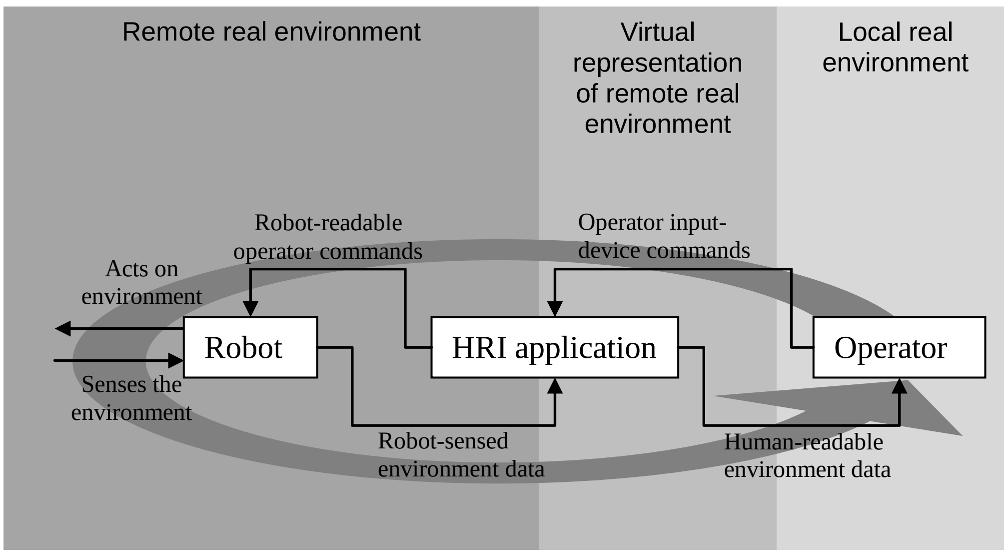


Figure 1: Simple representation of interaction cycle between robot and operator in an HRI system.

One problem in remotely operating and controlling a robot is being able to process all the data provided by the robot while performing mission-related tasks. The operator will be using the graphical robot interface not only as a means to understand the state of the robot and its surrounding environment, but also as a tool to complete mission goals. For urban search and rescue (USAR) tasks, this may be understood as a search task for victims after a catastrophic event. The excessive use of graphical interfaces without the correct fusion of elements can cause cognitive load on the operator and hence a decrease in the performance of the HRI system as a whole (Yanco, 2004; Drury, 2004; Neilsen *et al.*, 2006).

The purpose of this work is to evaluate how having part of the robot data be redirected to a vibro-tactile interface affects operator performance. Specifically, this project compares the effectiveness of two types of display interfaces for collision proximity feedback (CPF): a vibro-tactile one based on the TactaBelt (Lindeman *et al.*, 2006) and a graphical one which inherits some of the features of the Sensory Egosphere (Johnson *et al.*, 2003) and other proximity detection sensor interfaces (Yanco *et al.*, 2004; Yanco *et al.*, 2006).

The remainder of this paper is structured as follows. Section 2 provides an overview of the current work on HRI interfaces and how our work relates to it. Section 3 describes our interface. Section 4 presents the hypotheses for the user study that was carried out. The methodology used, from the definition of the virtual environment to the specification of the user study and experimental parameters is presented in section 5. Section 6 contains all the data results and analysis, which are discussed in detail in section 7. Last, conclusions can be found in section 8 followed by acknowledgements and references in sections 9 and 10 respectively.

Additionally, more information on the user study is provided in the appendices. The script used by the experimenter during the user study can be found in Appendix A. Appendix B contains the information contained in the user study instruction sheet. Last, Appendix C lists the questions contained in the user study post-questionnaire.

2. Related Work

Robot interfaces for USAR systems have been discussed and tested for more than a decade (Murphy *et al.*, 1996). Nevertheless, a standard model or a set of guidelines for HRI has not yet been widely accepted by the community.

Drury has pointed out a set of preliminary guidelines for HRI interfaces (Drury *et al.*, 2004). Scholtz has also specified a set of general recommendations for designing intelligent systems, but it may also be applied to HRI interfaces (Scholtz, 2002). Higher level interfaces for robot socialization and learning have also been a point of discussion and research in the community (Bien & Lee, 2007). Additionally, the AAAI competition has also been raising and identifying interesting issues about HRI interface design (Yanco *et al.*, 2004). Table 1 below encompasses categorized guidelines that are proposed by some of the above-mentioned research works.

Table 1: Guideline categorization according to the interface feature they are related to.

General guideline definition	Related guidelines
Cognitive load	• Lower cognitive load (Drury *et al.*, 2004); • Provide fused sensor information (Yanco *et al.*, 2004); • Minimize the use of multiple windows (Yanco *et al.*, 2004).
Situation awareness	• Enhance awareness (Drury *et al.*, 2004); • Provide a map of where the robot has been (Yanco *et al.*, 2004); • Provide more spatial information about the robot in the environment (Yanco *et al.*, 2004).
Efficiency	• Increase efficiency (Drury *et al.*, 2004); • Is the interaction language efficient for both the human and the intelligent system? (Scholtz, 2002) • Provide an interface supporting multiple robots (Yanco *et al.*, 2004).
Modality	• Provide help in choosing proper robot autonomy level (Drury *et al.*, 2004); • Is the necessary information present for the human to be able to determine that an intervention is needed? (Scholtz, 2002) • Provide robot help in deciding which level of autonomy is most useful (Yanco *et al.*, 2004).
Scalability, Portability & Sociability	• Does the interaction architecture scale to multiple platforms and interactions? (Scholtz, 2002) • Design should be a social activity and robots should be social agents (Bien & Lee, 2007).
Reuse	• Does the interaction architecture support evolution of platforms? (Scholtz, 2002) • All applications should be reapplications (Bien & Lee, 2007).
Robustness	• Must tolerate ambiguity (Bien & Lee, 2007).

An attempt to categorize the interface features currently suggested and used in USAR interfaces led us to the results presented in Table 2. An increasingly-ordered histogram table on different interface features for six HRI interfaces from a variety of research projects (Johnson *et al.*, 2003; Nielsen *et al.*, 2006; Nielsen *et al.*, 2007; Yanco *et al.*, 2004, Yanco *et al.*, 2006; Drury *et al.*, 2004) is presented.

The information in this table describes the features that are potentially more reusable for different tasks over other ones that are more specific to a certain robot or mission. It is obvious that video feedback is a necessary feature for all interfaces. Notice however, that other features, such as a map, robot pose and orientation, and sensor monitoring, are also of relevance for most interfaces. This indicates that some features should perhaps be considered as general requirements of HRI interfaces. Assuming this as a premise, a standard configuration for interaction with such features of interfaces could be defined.

Table 2: Summary of interface features and their occurrences in HRI projects.

Feature	**Number of occurrences**
One or more video feeds	6
Map	5
Orientation	5
Pose	4
Sensor monitoring	4
Directional event indicator	3
Option selectors: operation mode, camera, action	2
Waypoint or landmark tagging	2
Camera configuration feedback	2
Incremental map building algorithm	1
3rd person view	1

By relating the interface features to the general proposed HRI guidelines, as presented in Table 3, it is noteworthy that the focus of most projects is on improving situation awareness and efficiency, but not on reducing cognitive load and creating systems that are scalable, portable, reusable or robust. Note that some of the features could improve more than one parameter. For example, one-video with more emphasis generally reduces cognitive load compared to having two videos of the same side (Yanco *et al.*, 2006), but one could argue that having video may also affect situation awareness, for 1 single larger video window may be a better source of SA than two smaller ones.

The interface proposed here is aimed to change this HRI interface development trend by proposing an interface that will attack the problem of the operator's cognitive load by using feedback devices in addition to graphical ones.

Table 3: Categorization of currently implemented interface features according to general HRI guidelines.

Interface feature	Improvement						
	Cognitive load	Situation awareness	Efficiency	Modality	Scalability, Portability & Sociability	Reuse	Robustness
One-video with more emphasis instead of more videos with the same emphasis	X						
Allow multiple views		X					
Provide orientation of robot and camera		X					
Provide robot pose		X					
Provide directional alerts for blockage or collision		X					
Provide sensor information		X					
Provide camera zoom information		X					
Provide operation mode information		X					
Provide orientation for sensor events		X					
Environment auto-mapping mechanism		X					
3rd person view		X					
Enable placement of POIs or waypoints			X				
Integrate collision avoidance mechanism			X				
Autonomous actions preprogrammed (escape, pursuit)			X				
Automatic direction reversal			X				
Multiple video windows			X				
Provide different teleoperation modalities				X			
Our Interface Novel Feature							
TactaBelt as channel for sensor information	X	X					
TactaBelt as channel for directional information	X	X					

However, one can also think of many other types of input and output (I/O) devices that are still barely being explored in HRI, such as spatial audio, low-resolution stereoscopic displays and using operator's body movement to accomplish tasks in the remote environment. Many of these are can be derived from Virtual Reality I/O techniques (Bowman *et al.*, 2005) and integrated with current HRI sensor technologies.

3. Our Interface

We have designed an interface that attempts to follow a superset of the guidelines that have been proposed in the field. It also attempts to merge good features from interfaces previously proposed and tested by other research groups.

Our design uses as a starting point the interface proposed by Nielsen (Nielsen *et al.*, 2006; Nielsen *et al.*, 2007). The operator is presented with a third-person view of a 3D virtual representation of the robot, called *avatar*, along with data collected by the robot as seen in Figure 2. These data include a video feed from the real environment, sensor data about the location of object surfaces near the robot, and potential collision locations.

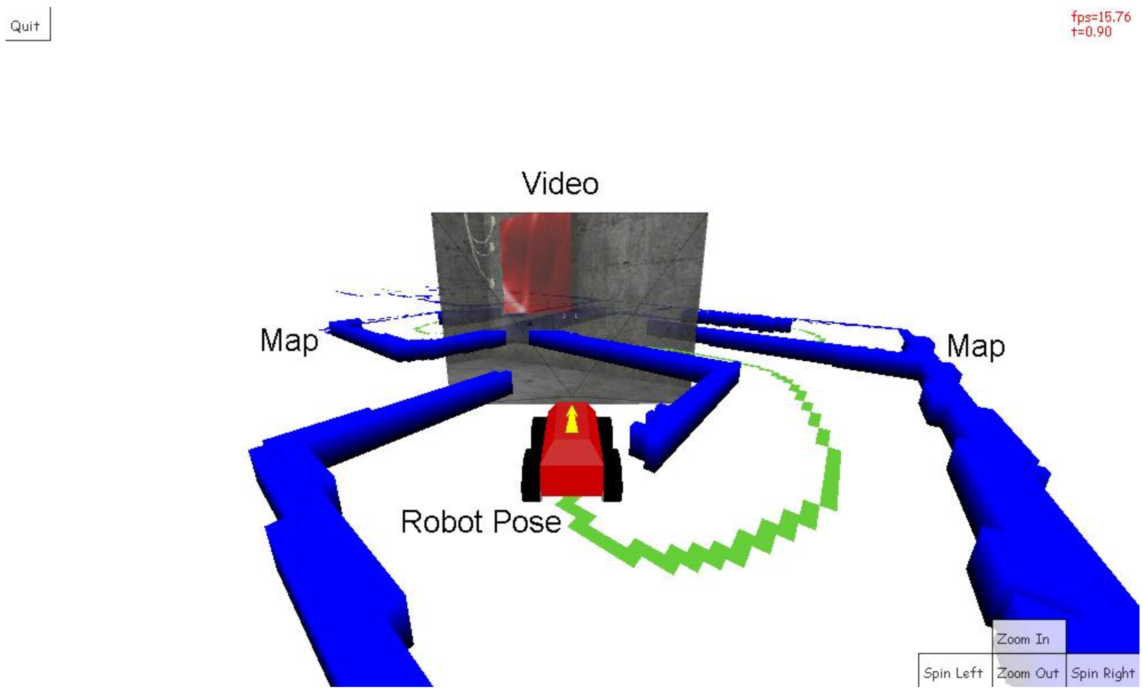


Figure 2: Nielsen's HRI interface for robot teleoperation (Nielsen *et al.*, 2006).

In our interface, a panel located in front of the robot avatar projects data from the robot's video camera. The camera, and hence the panel, can be rotated about both the vertical and horizontal axes, up to an angle of 100 degrees horizontally for each side and 45 degrees vertically in both upward and downward directions, and relative to the robot initial orientation.

A ring surrounding the robot avatar indicates imminent collisions near the robot, similar to the Sensory EgoSphere proposed by Johnson (Johnson *et al.*, 2003) but with a more specific purpose. The brighter the red color in the ring the closer to a collision point the robot is.

The same type of feedback is also provided as vibration through the vibro-tactile interface, the TactaBelt. The latter consists of eight pager motors, also called tactors, arranged in a ring around the torso, with the motors spaced evenly (see Figure 4b). The more intense a tactor in the TactaBelt vibrates, the closer the robot is to colliding in a certain direction.

Following the experimental results and conclusions by Nielsen (Nielsen *et al.*, 2006), the map of the environment was set to be represented in a projected manner on the ground in the form of blue lines as the robot captures data from the environment. The robot avatar is placed on map where the robot is physically located in the real world.

Figure 3 illustrates the main components of the graphical interface. Blue lines are printed on the ground and represent object surfaces that are sensed by the robot as it passes close to them. Previously sensed walls and their locations are stored and displayed. For the example in Figure 3, the robot passed close to a large number of the objects in the scene before and thus most of the walls in the environment can be seen from a distance. The eight cylindrical components of the graphical ring are presented in different red saturation levels around the robot avatar, in front of which the camera panel can also be seen. The panel rotations occur relative to the robot avatar and match the real robot camera rotations controlled by operator input. A chronometer is presented in the top right hand corner of the screen. It is triggered once the training session finishes and the robot is transferred to the virtual environment (VE) where the actual experiment takes place.

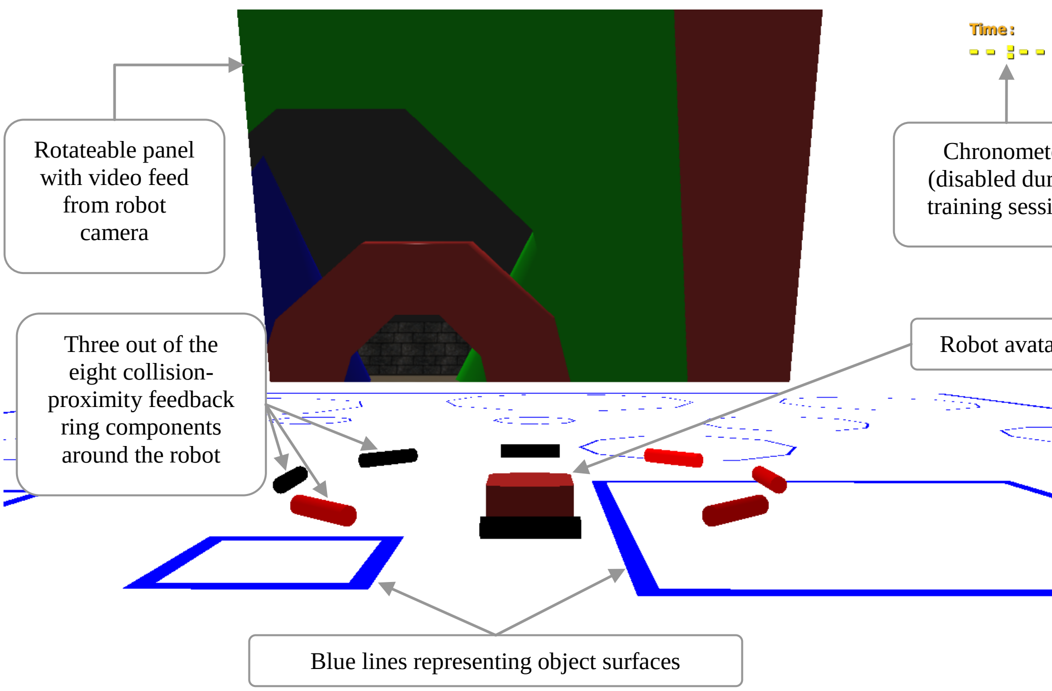


Figure 3: Main Components of the graphical interface.

The controller used in the experiment was a Sony PlayStation Dual-shock 2 (Figure 4a). The TactaBelt was custom made, and consists of a set of eight tactors located at the cardinal and intermediate compass points, with forward being north (Figure 4b). Each tactor is a simple DC motor with an eccentric mass, similar to those used in mobile phones. The amount of current controls the level of vibration for each tactor. This control is possible through the interface API and hardware provided by the TactaBox (Lindeman *et al.*, 2006).

(a) (b)

Figure 4: Interface used in addition to the standard computer monitor: (a) PlayStation 2 dual-shock controller; (b) TactaBelt.

The machine used for running the experiment was a DELL XPS 600 with 2 GB RAM and a Pentium (R) D Dual-core 3GHz processor. The environment was run in a window with resolution of 1280x1024 at an average frame rate of 30 frames-per-second (fps). The graphics card used was a GeForce 7800 GTX with 256MB of memory.

4. Hypothesis

We hypothesize that the use of vibro-tactile feedback will reduce operator cognitive load when compared to using only a graphical interface. The hypothesis will be validated by an evaluation of user performance on a simple object-search task performed by subjects in a user study. The reduction in cognitive load should be indirectly perceived by measuring dependent variables on the study such as the number of searched objects that are found, number of robot collisions with the environment, task time and spatial understanding of the environment, as well as identification of the locations of objects.

The use of the graphical ring, as well as the TactaBelt, should cause an improvement in subjects' perception of the surrounding environment, which would be an indication of an increase in their situation awareness level. This improvement should be especially evident through a reduction in the number of collisions, but also in other experimental variables. By enhancing the navigational feedback and making it more intuitive (directional feedback) and less visual (vibro-tactile feedback), subjects using the enhanced interfaces will be able to focus more on the task and, hence, find a larger number of objects. Therefore, task time, number of collisions, and number of objects found are the most important measurements to validate or reject our hypothesis, which is split into the following two more-specific hypotheses.

Our first hypothesis is that **subjects using either the TactaBelt or the graphical ring feedback interfaces should have an increase in navigational performance measured by a reduction in the number of collisions and hence in the time taken to perform the task in relation to the control group**, that is using neither interface. Additionally, these subjects will be able to focus their attention more on the task itself rather than on navigation. As a result, **there should also be an increase in the number of objects found by these subjects compared to the control group, and a better reporting on the location of the objects and understanding of the environment** as measured using post-test sketchmaps.

Our second hypothesis is that **subjects who are using both the TactaBelt and the graphical ring feedback interfaces should have an even larger increase in performance measured by a reduction in the number of collisions and in the time taken to perform the task** compared to all other groups. Moreover, **an increase in the number of objects found and more-accurate sketchmaps being drawn should also be evidenced** compared to all other groups of subjects.

5. Methodology

As already mentioned, a user study was carried out to confirm our hypotheses that the use of either or both feedback interfaces would imply in an improvement in operator performance and reduction of cognitive load measured in terms of number of collisions, task time, number of objects found and spatial understanding of the environment. The user study is described in this section.

5.1. User Study

The task that users had to complete consisted of locating red spheres in the ruins of a small closed environment. A total of nine spheres were hidden in the environment. The users did not know in advance the number of spheres hidden. After the experiment was over, users were asked to sketch a map of the task space and the approximate location of the spheres.

The user study consisted of a between-subjects experiment. The independent variable was the type of collision-proximity feedback (CPF) interface. Subjects were divided into four groups: the first group ("None") would operate the robot without having collision-proximity feedback from either the graphical ring or the TactaBelt. The second ("Ring") would receive feedback from the graphical ring. The third ("TactaBelt") would receive feedback from the TactaBelt. The fourth and last ("Both") would receive feedback from both the graphical ring and TactaBelt.

Subjects could control the robot and its camera using the two analog joysticks of the gamepad. Two trigger buttons from the gamepad were also used to allow subjects to take pictures of the environment. These pictures were useful for subjects in the map sketching task during the post-questionnaires session as further explained below.

The dependent variables were the number of collisions during the task, the time taken to accomplish the task, the number of spheres found, and the accuracy of the sketchmaps.

The user study can be summarized by a list of eight steps for each subject, some of them further explained in the paragraphs following this list.

1. The subject read and signed an IRB-approved consent form for the experiment;
2. Demographic information was collected from the subject;
3. The subject read a sheet of paper with instructions for the experiment and asked any questions about the experiment;
4. The experimenter explained how to control the robot and its camera using the controls on the gamepad;
5. The experimenter explained about the training session and then started it when the subject was ready;
6. Once the training session was finished, the experimenter explained that the subject would be moved to the real task now and briefly reviewed the task to be accomplished. The experiment started when the subject was ready;
7. During the main experiment, the experimenter took general notes about the subject and his performance during the experiment;
8. Once the main experiment task was over, the subject filled in a post-questionnaire containing the sketchmap and experiment feedback information.

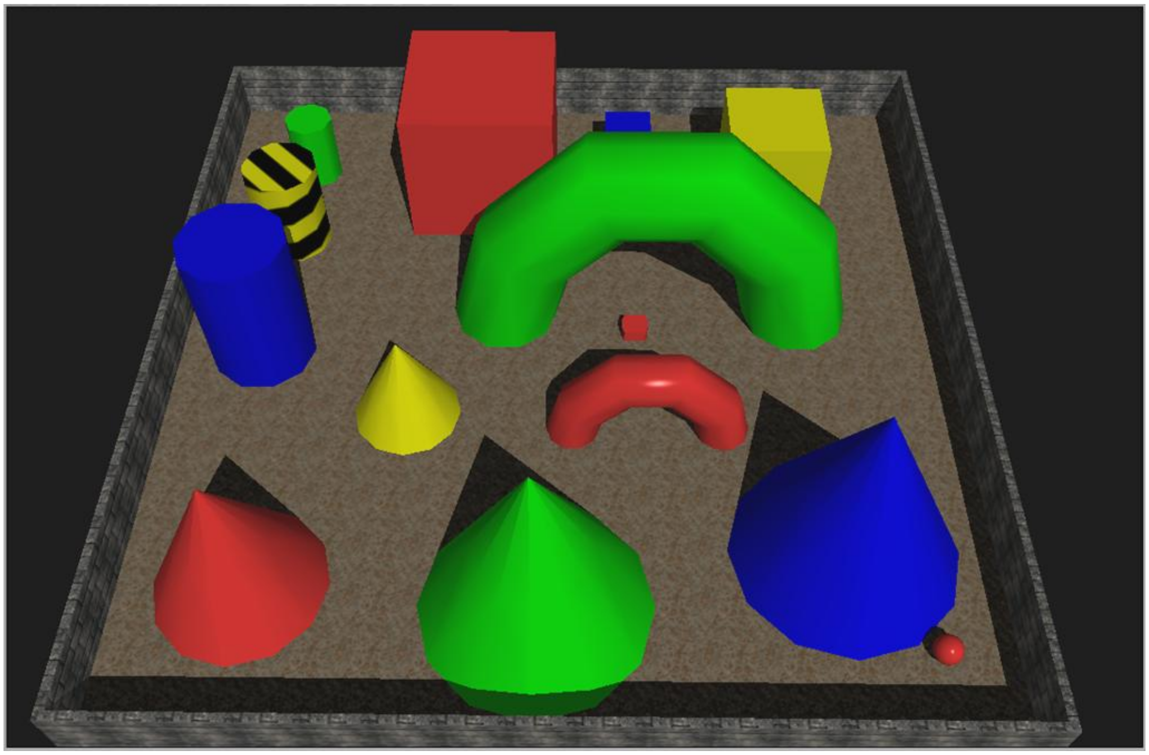
A description of the script used by the experimenter as a guideline during the experiment can be found in Appendix A.

The demographic questionnaire collected subject information about their gender, age, how often they played video-games and used or worked with robots. For the last two questions, the possible answers were: “daily” (1), “weekly” (2), “seldom” (3) or “never” (4).

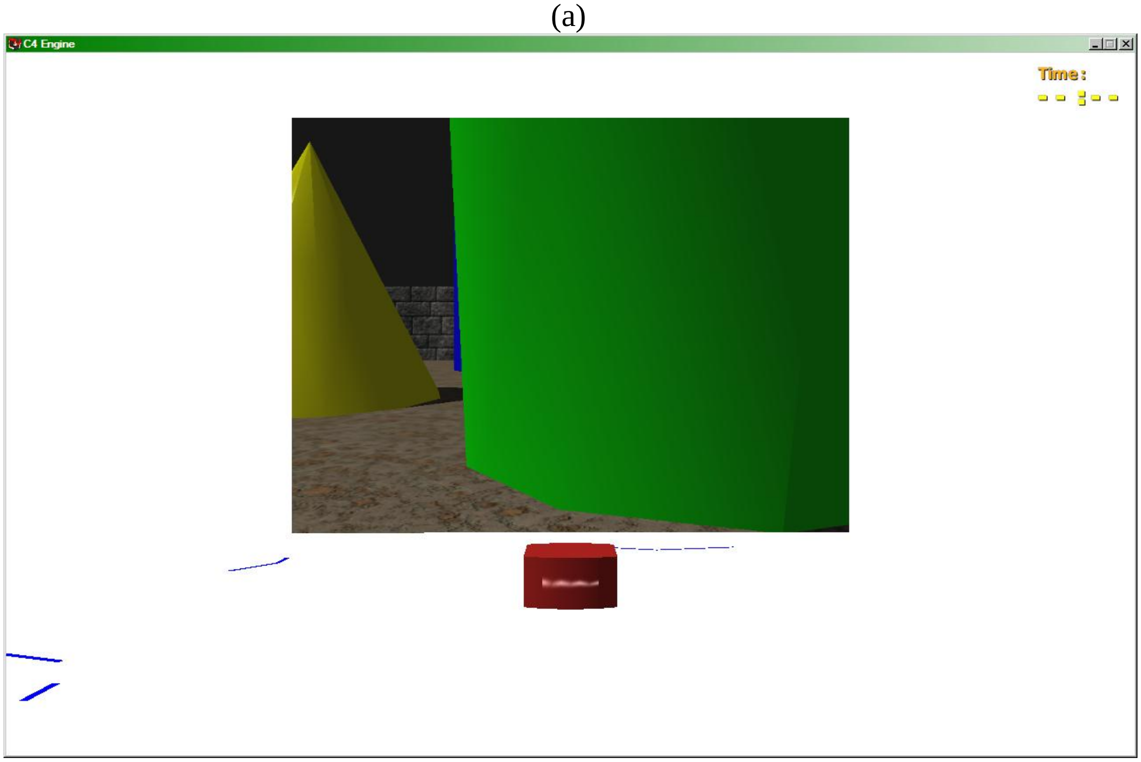
The instructions, presented in Appendix B, consisted of a single page of information containing the description of the experiment, the task that should be completed, the interface, and how subjects should behave before, during, and after the experiment.

The training session happened in a virtual training room larger than the one used during the real task session. While the virtual house where the experiment took place was 8m in width by 10m in length, the training session room was 15m in width by 15m in length. Additionally, the training room was an open space while the experiment room had many obstacles the user needed to maneuver around.

The training room contained a set of large geometric primitives colored with basic colors. A single red sphere was hidden behind one of these primitives. The task in the training room was to find the hidden red sphere and take a picture of it. This gave the subject time to practice and get used to the robot controls. During the training session, if subjects seemed to be already comfortable with the robot controls but were having problems in finding the red sphere, the experimenter would intervene and give them hints on the location of the sphere so that they could practice taking pictures, ask questions and then move on to the real experiment. Figure 5 gives an overview of the training room.



(a)



(b)

Figure 5: Training environment from a bird's eye view (a) and from the operator's perspective (b).

Data on place and time of the collisions were recorded as well as the time spent in performing the task. Additionally, the periods of time spent during the training session and sketching the location of the spheres were recorded for some of the subjects; we didn't think to collect such data until half-way through the subjects.

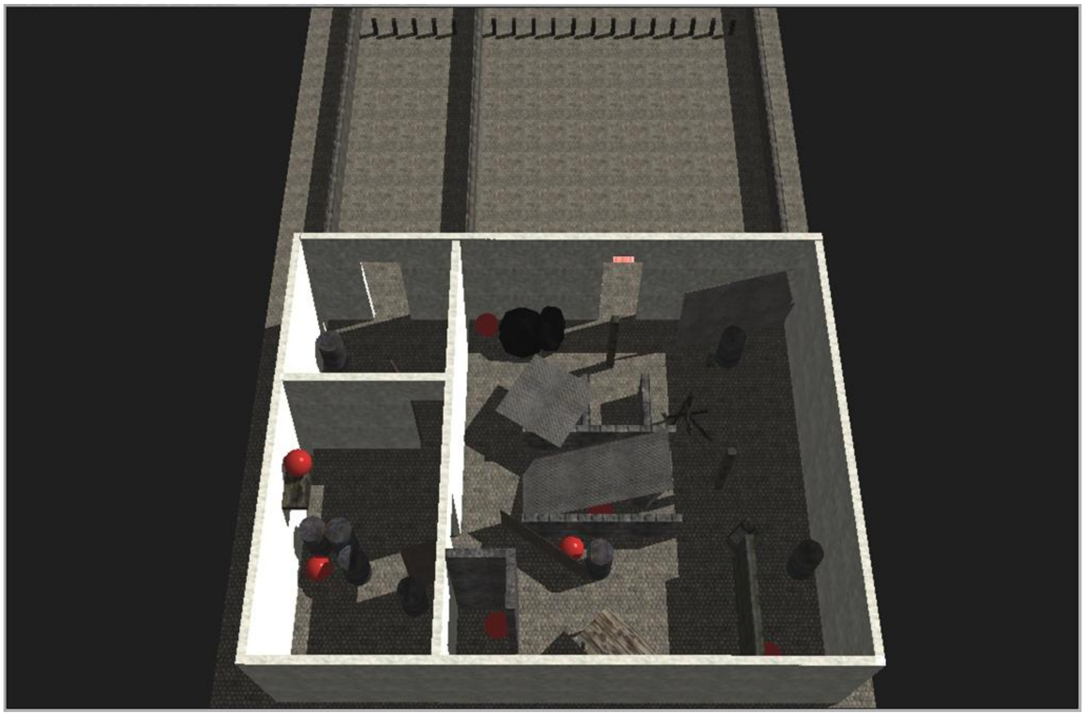
A post-questionnaire asked subjects to report the number of spheres found and their location by sketching a map of the environment. They were provided with the pictures they took during their traversal of the environment to help them in sketching. The images were displayed on a Web page presented at 800×640 resolution. The top of such a Web page is presented in Figure 6.

Figure 6: Web page example used during map sketching.

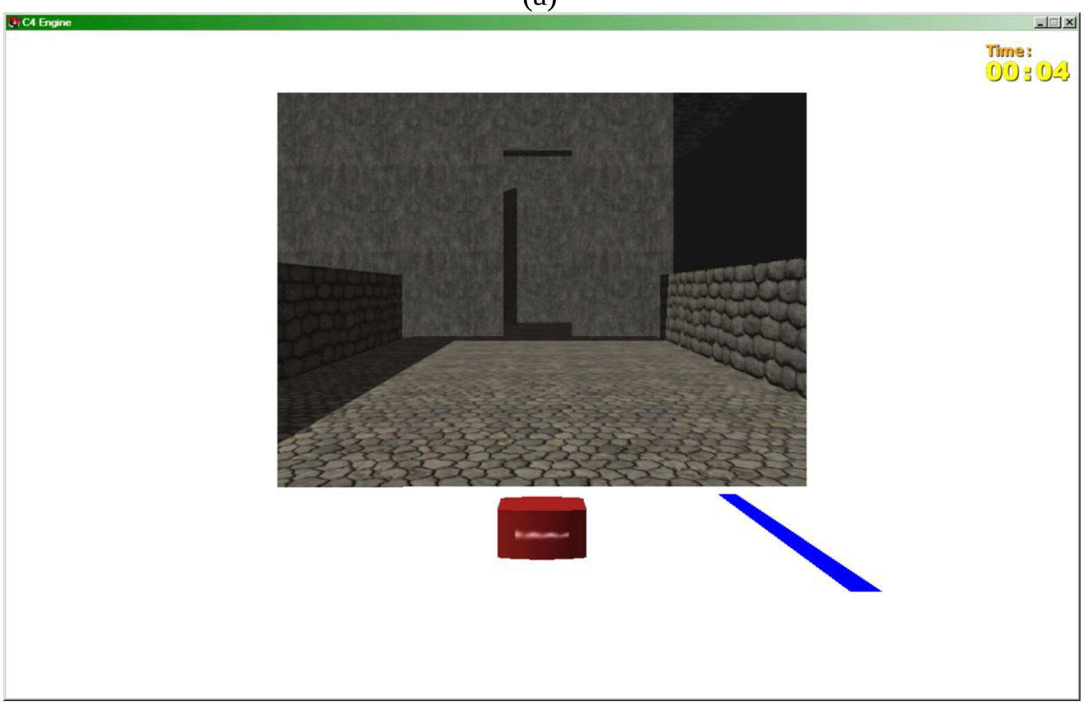
The sketchmaps were evaluated following criteria proposed by Billinghurst & Weghorst (1995). The first criterion was map goodness, which was evaluated on a scale from 1 to 5, instead of the original scale from 1 to 3. The criterion for grading map goodness was how well the sketchmap would help in guiding someone through the environment. The second criterion was counting the number of objects of different classes or groups that were drawn. The objects were divided into three groups: walls, doorways, and debris. These groups were graded separately. Each object found corresponded to an increment of one point to their object group grade. The third and last criterion was a general grading and analysis of the correct placement of objects relative to other nearby objects. Sample sketchmaps are presented in Figure 33. Spheres were not considered, since pictures would allow subjects to position them correctly relative to nearby objects most of the time.

5.2. Virtual Environment

The robot side of the system was simulated using a virtual environment (VE). In fact, two VEs, constructed using the C4 game engine (www.terathon.com), were used in our application. The first described the real world where the robot was inserted and was supposed to complete a task. The second represented the robot teleoperator interface as seen from the point of view of the operator. The environment where the search task took place is presented in Figure 7a, and the operator view is shown in Figure 7b.



(a)



(b)

Figure 7: Experimental virtual environment in C4 engine (a) from a bird's eye view and (b) from the operator's perspective. This version does not have the graphical ring and may represent the view of a subject in the "None" or "TactaBelt" group.

6. Results

The results are presented in eight parts. The first part contains demographics about the population. The second attempts to validate the homogeneity of the groups of subjects according to gender, age, and experience, and how the lack of experience may affect the results. The third part presents general results about the data collected in the experiment. The fourth part goes into more detail about the results presented in part 3 by comparing data from different groups. The fifth attempts to identify possible correlations in different pairs of experimental factors. The sixth part reports a summary of the notes taken for all subjects while the experiment was being run. The seventh one presents results related to the maps sketched by subjects. The eighth and last part summarizes subjects' comments on post-questionnaires.

6.1. Population Information

The population consisted of university students. A total of 13 females and 14 males participated in the user study. The mean age of the population was of 20 years, six months, with standard deviation of 5 years, three months, and median of 19 years. The age distribution between genders was also reasonably homogeneous as can be seen in Figure 8.

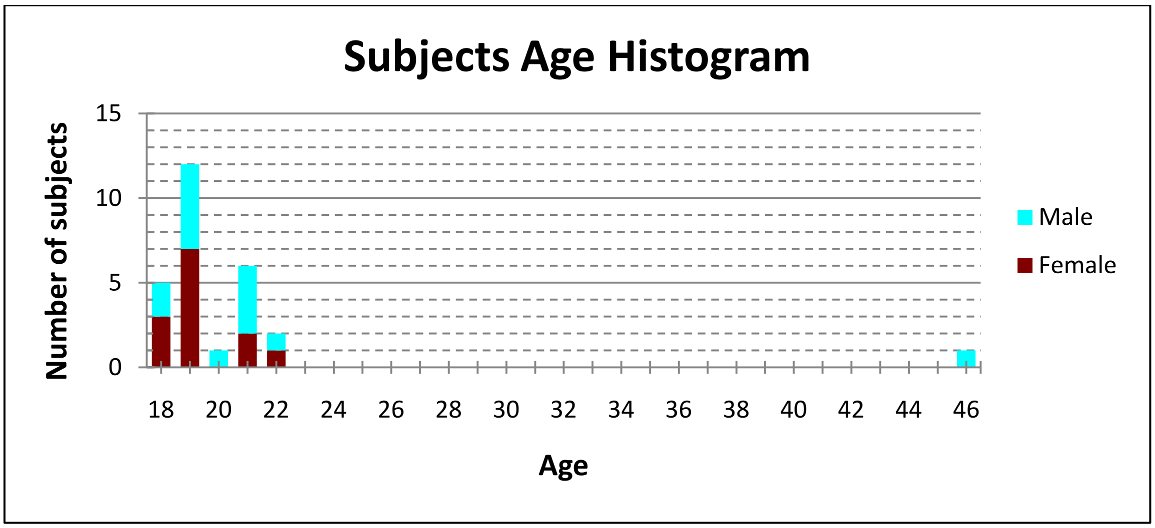


Figure 8: Histogram for subject age and the distribution between genders.

Most subjects reported playing videogames only seldom. Considering a scale from 1 to 4, where 1 is associated with daily, 2 with often, 3 with seldom and 4 with never playing videogames, results indicate a mean of 2.81 with a standard deviation of 0.96 and a median of 3. The histogram for subjects' videogame experience is presented in Figure 9.

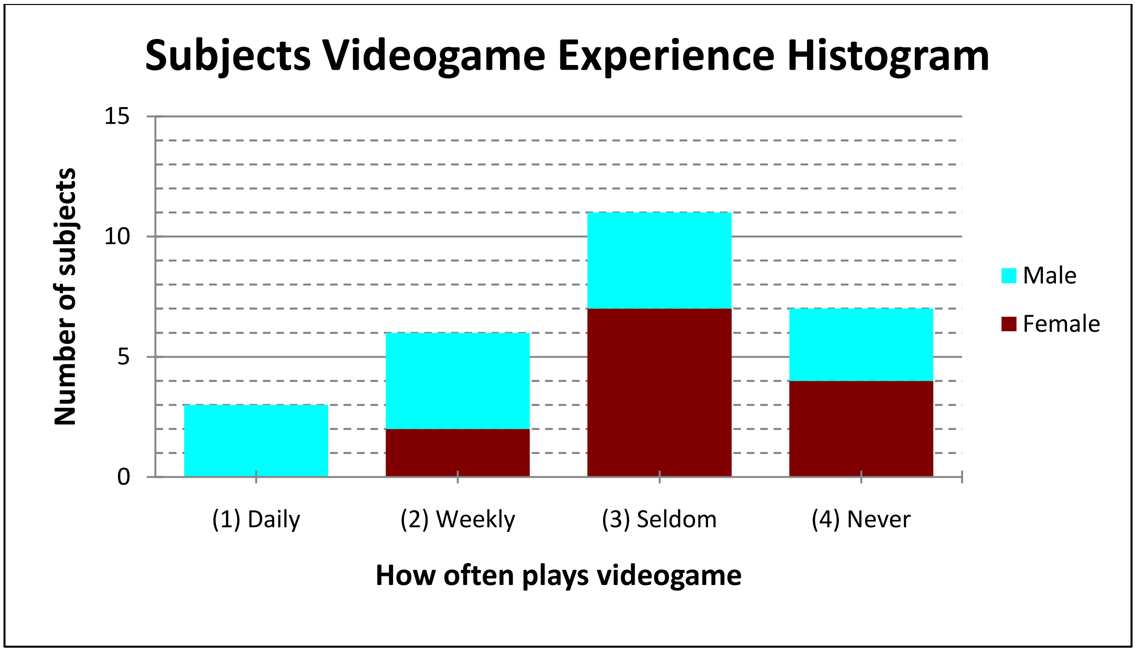


Figure 9: Histogram for subjects' videogame experience and its distribution between genders.

Most subjects reported they had never used or worked with robots before. Some male subjects had some experience with robots by either participating in high-school robot competitions or taking a course related to the subject. Considering the same scale from 1 to 4 used for videogame experience, results indicate a mean of 3.81 with standard deviation of 0.39 and median of 4. The histogram for subject robot experience is presented in Figure 10.

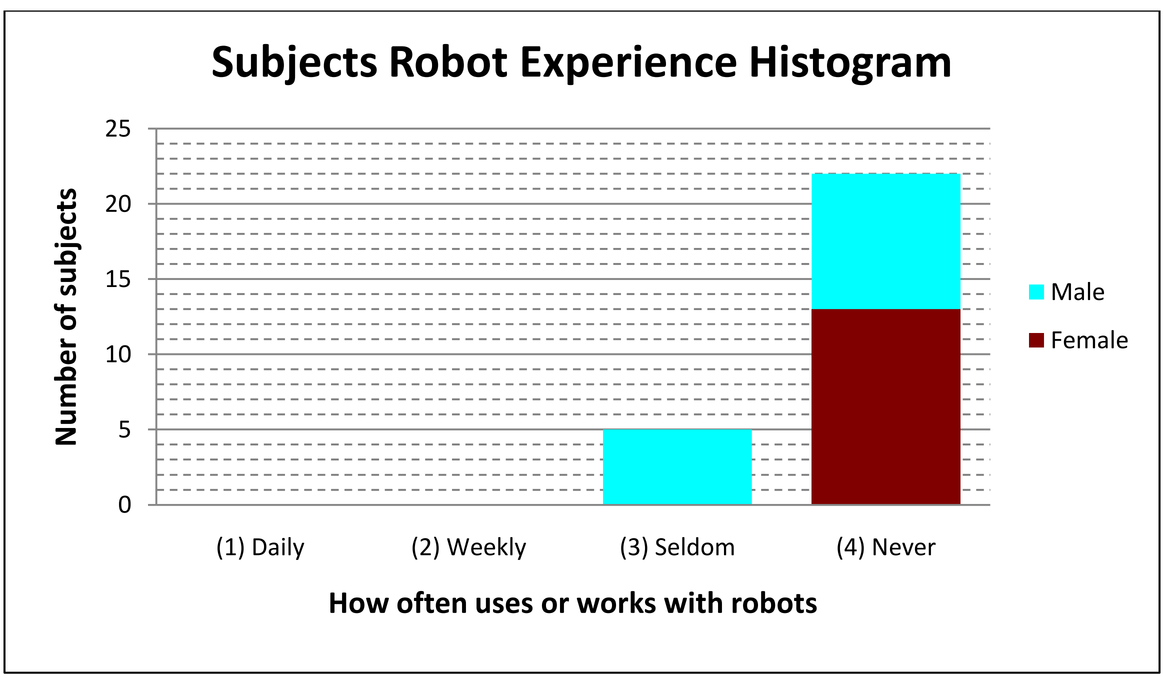


Figure 10: Histogram for subjects' robot experience and its distribution between genders.

6.2. Homogeneity Check

In order to ensure our results are not due to inconsistency among groups, analyses of the distribution homogeneity amongst groups of subjects according to experimental variables were conducted.

Gender Distribution

An analysis on how genders were distributed amongst groups (Figure 11) shows that each group had at least two members of each gender. The mean number of females per group was of 3 with a standard deviation of 1 and median of 3. The mean number of males per group was of 4 with standard deviation of 1 and median of 4.

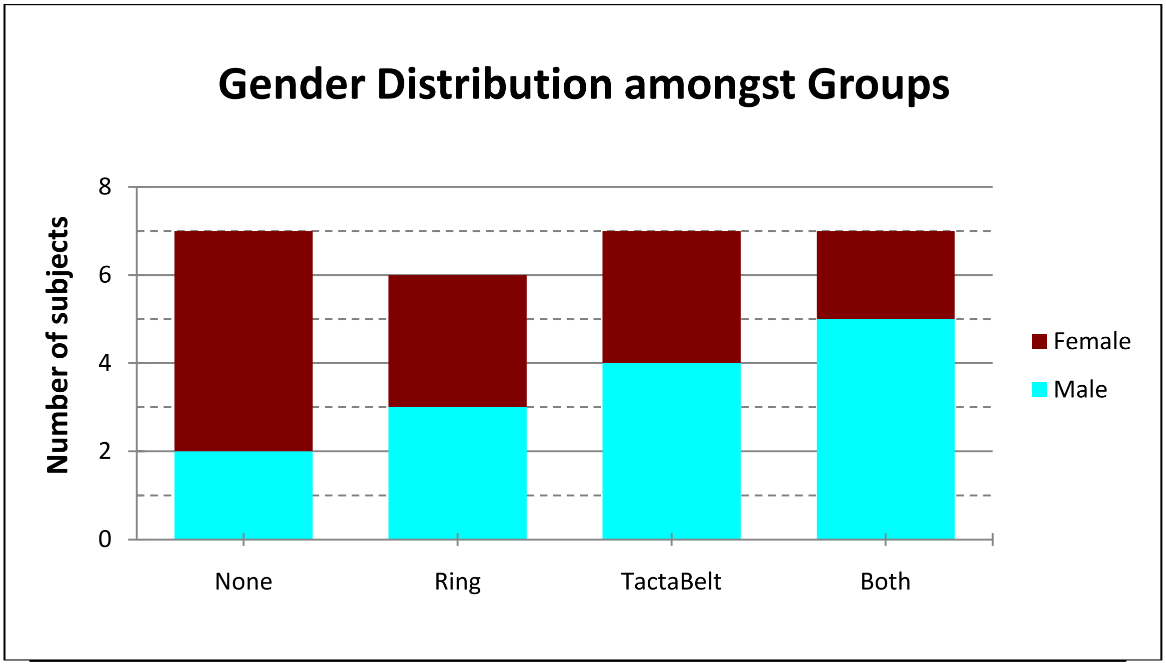


Figure 11: Gender distribution amongst groups.

An analysis of how age was distributed amongst groups (Figure 12) shows that members' age for each group seems to be mostly concentrated around the late teens. The mean age, standard deviation and median for each group as reported in Table 4 confirms this claim.

Looking at the median values, only group "Both" has a value that significantly deviates from the late teenage estimation. This group was mostly composed of fourth-year students.

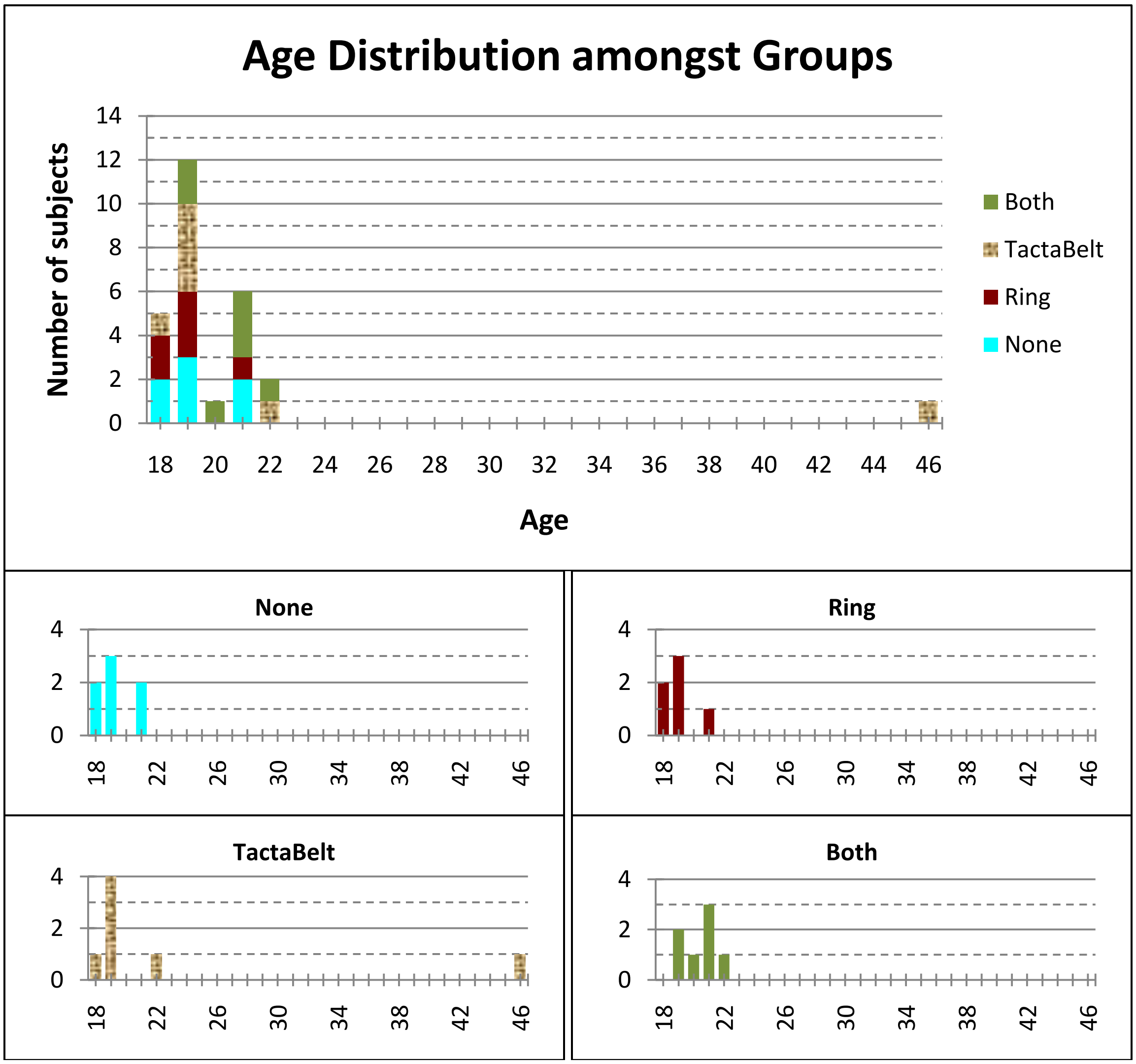


Figure 12: Age distribution amongst different interface groups.

Table 4: Mean, standard deviation and median of subject age by group.

	Group "None"	Group "Ring"	Group "TactaBelt"	Group "Both"
Mean	19.3	19.0	23.1	20.4
Standard deviation	1.25	1.09	10.16	1.13
Median	19	19	19	21

We considered subjects' videogame experience by group. According to the results presented in

Table 5 and Figure 13, groups "None" and "TactaBelt" had similar levels of experience. Group "Ring" seems to have been a less experienced group, while group "Both" seems to have been a more experienced one.

Table 5: Mean, standard deviation and median of subjects' videogame experience by group.

	Group "None"	Group "Ring"	Group "TactaBelt"	Group "Both"
Mean	2.9	3.3	3.0	2.1
Standard deviation	1.07	0.82	1.00	0.69
Median	3	3	3	2

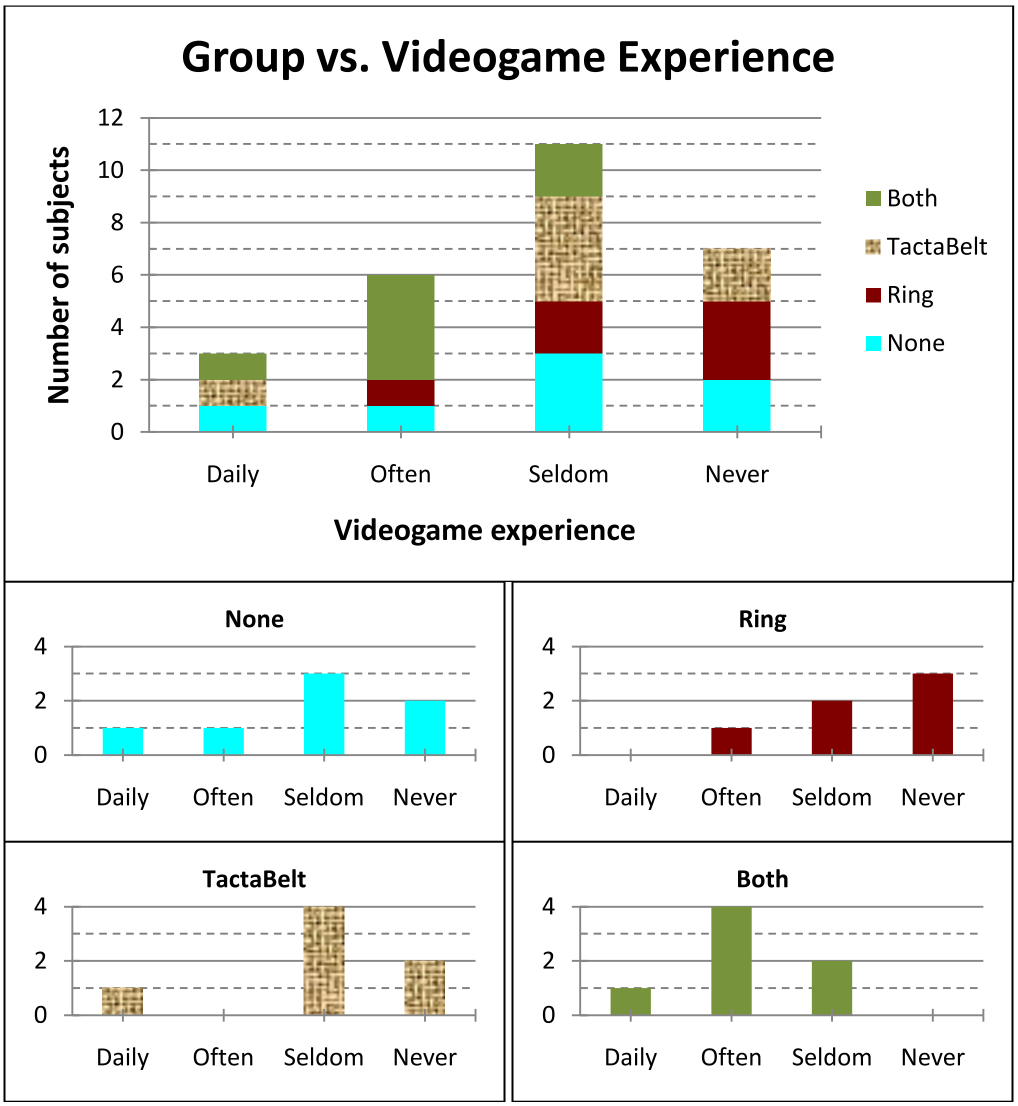


Figure 13: Videogame experience distribution amongst groups.

We also analyzed group homogeneity in terms of experience with robots. In fact, few subjects had experience with robots. These subjects were part of only groups “Ring” and “Both”, as viewed in Table 6 and Figure 14.

Table 6: Mean, standard deviation and median of subjects’ robot experience are presented separately for each group.

	Group “None”	Group “Ring”	Group “TactaBelt”	Group “Both”
Mean	4.0	3.5	4.0	3.7
Standard deviation	0.00	0.55	0.00	0.49
Median	4	4	4	4

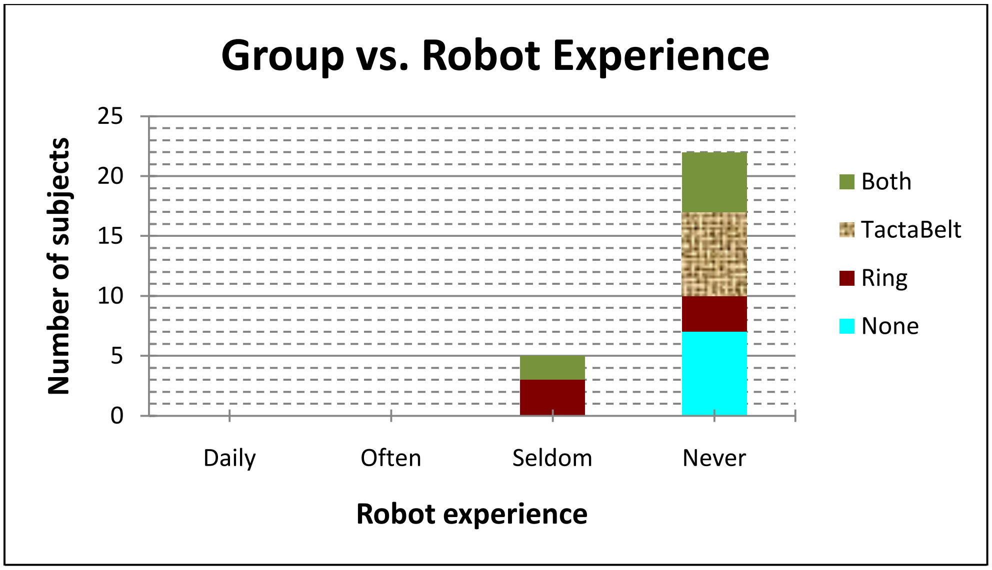


Figure 14: Robot experience distribution amongst groups.

6.3. General Results

The general results for task-completion time are presented in Figure 15 and Table 7, the latter also including the results for number of collisions. Task-completion time data appears to be distributed in a bimodal fashion with peaks at around 4 and 9 minutes.

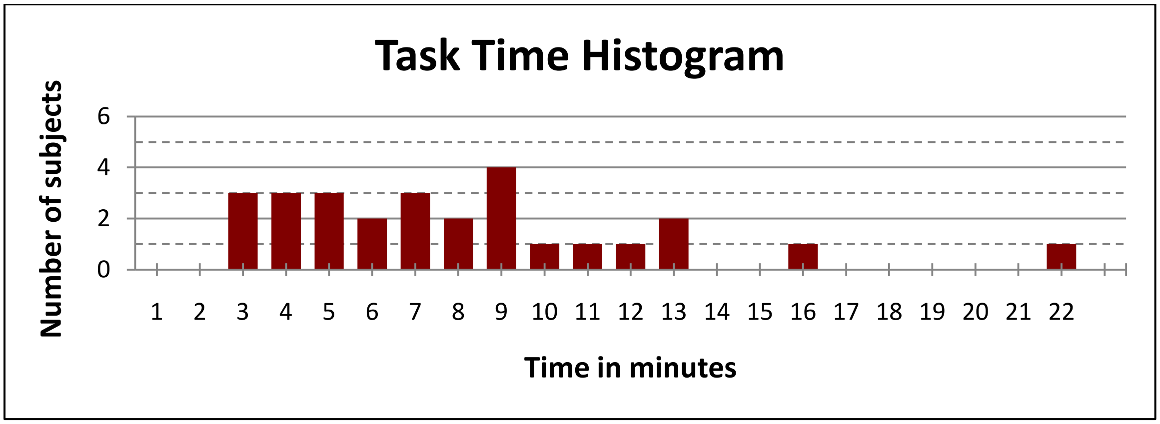


Figure 15: Frequency histogram for time spent in performing the required task.

The mean, standard deviation, and median for the number of collisions per subject are presented in Table 7, and the histogram is presented in Figure 16.

Table 7: Mean, standard deviation and median of subjects' task-completion time and number of collisions.

	Task Time (hh:mm:ss)	Number of Collisions
Mean	485 s (00:08:05)	36
Standard deviation	263 s (00:04:23)	34
Median	448 s (00:07:28)	23

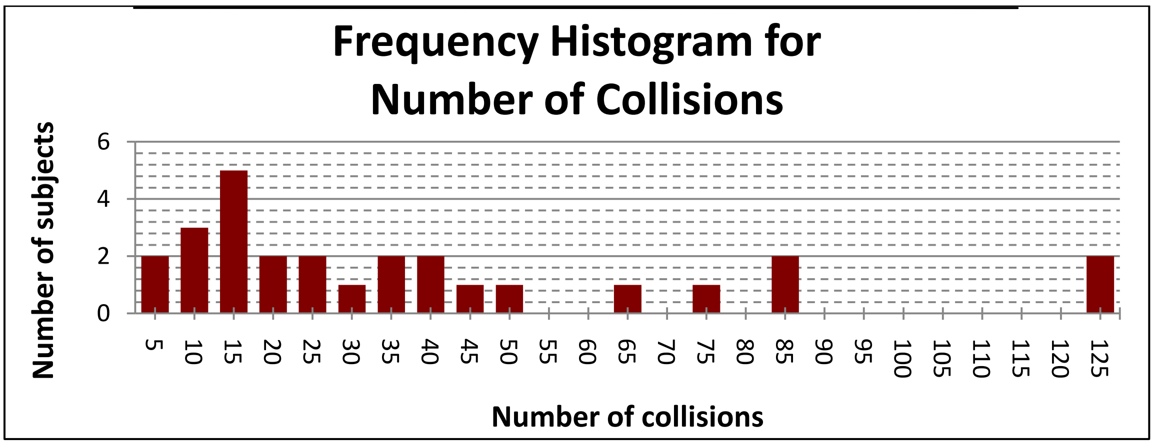


Figure 16: Frequency histogram for number of collisions between robot and objects in the environment while performing task.

The amount of time spent in the training session and map sketching were recorded for 15 subjects. The mean training session time was of 00:03:58 with standard deviation of 58 seconds and median of 00:03:50. The mean map sketching time was of 00:09:30 with standard deviation of 00:05:02 and median of 00:08:20. The histograms for both results are presented in Figure 17 and Figure 18. In Figure 17, only four columns were plotted because all subjects had a training time between 3 and 6 minutes.

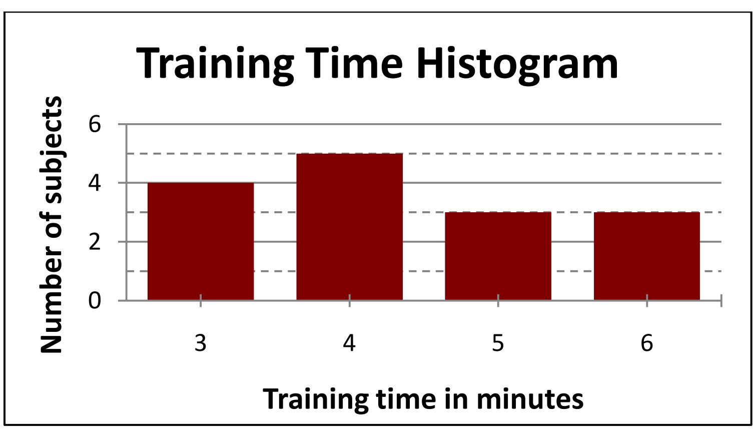


Figure 17: Histogram for time spent in training room.

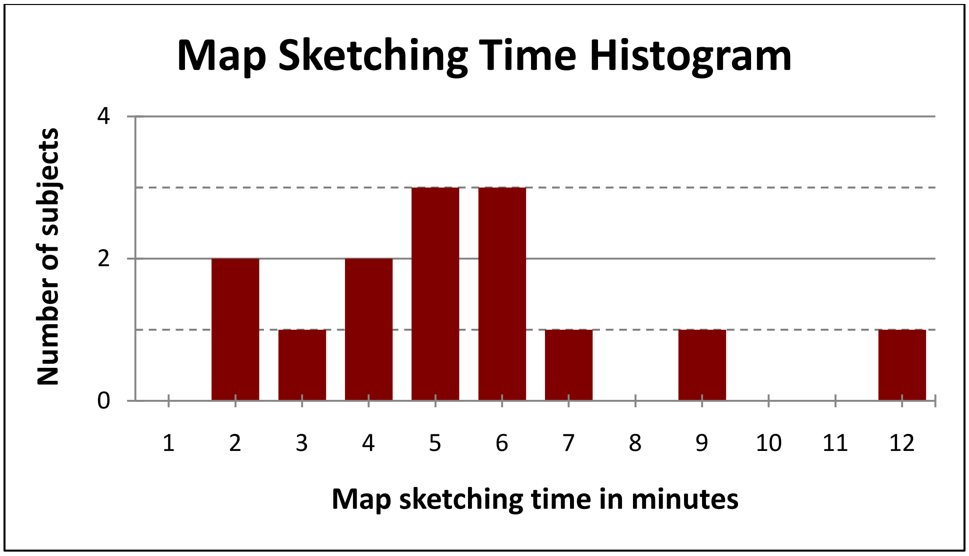


Figure 18: Histogram for time spent in map sketching.

The histogram for the number of spheres found, among the total of 9 spheres that were hidden, is presented in Figure 19. The mean number of spheres found was 4.48 with a standard deviation of 2.21 and a median value of 4.

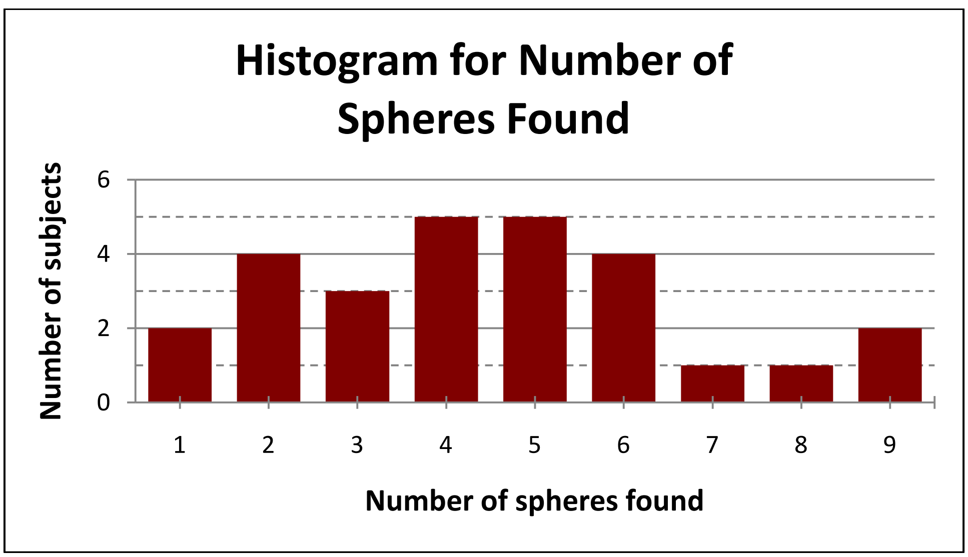
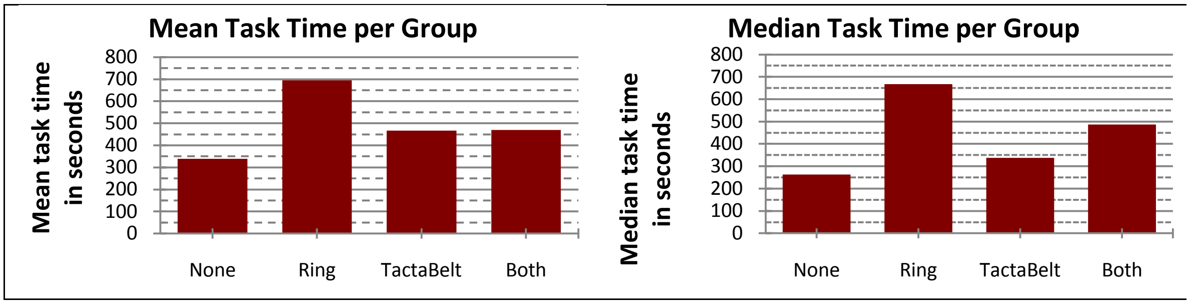


Figure 19: Histogram for number of spheres found.

6.4. Performance Comparison between Groups

This section compares the above results on a per-group basis to identify the effects the different CPF (collision-proximity feedback) interfaces had on task performance. For these and all other comparisons we used a single-factor ANOVAs with confidence level of p=0.05.

A comparison on task time between groups led to no statistically significant differences (F=2.319, $F_{critical}$=3.028, p=0.102, p = 0.05). The mean and median task time per group are presented in Figure 20.



(a) (b)

Figure 20: Mean (a) and median (b) task time per group.

The task time histogram previously presented in Figure 15 is now presented in Figure 21 with subjects from different groups being identified and separated. Notice the dispersion of group "Ring" results and the fluctuation of the other three groups' results between 4 to 5 minutes and 7 to 9 minutes.

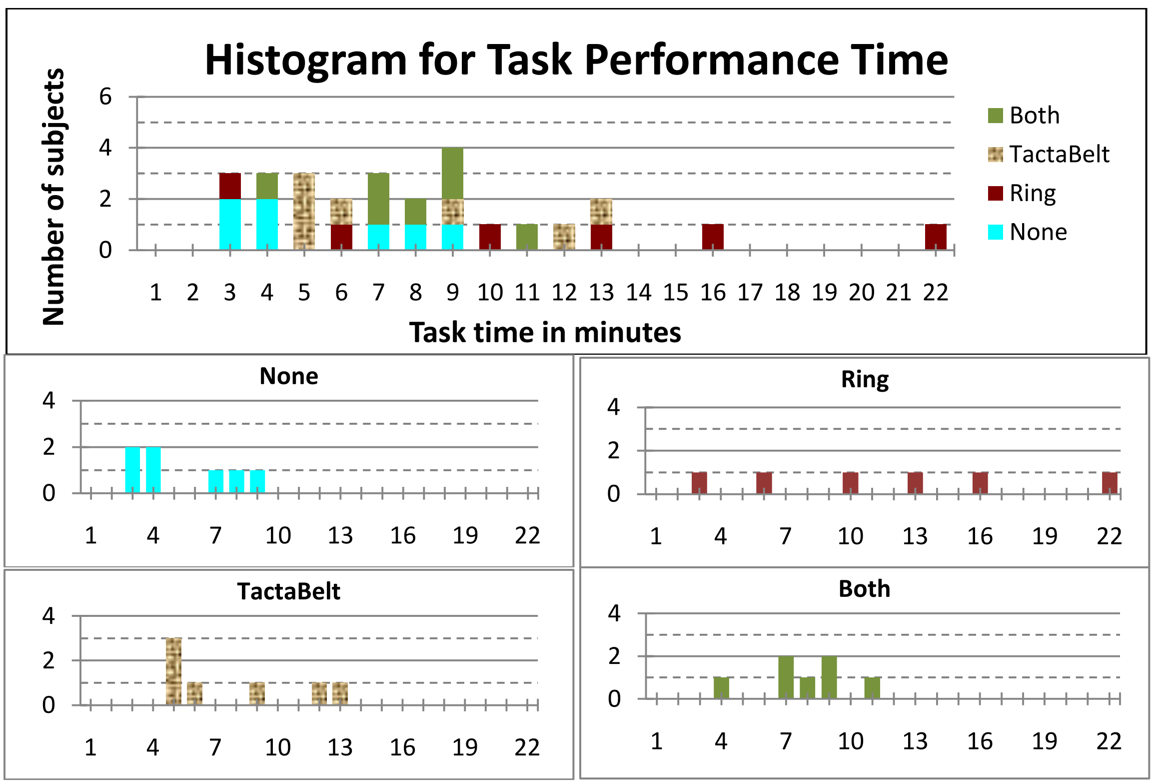
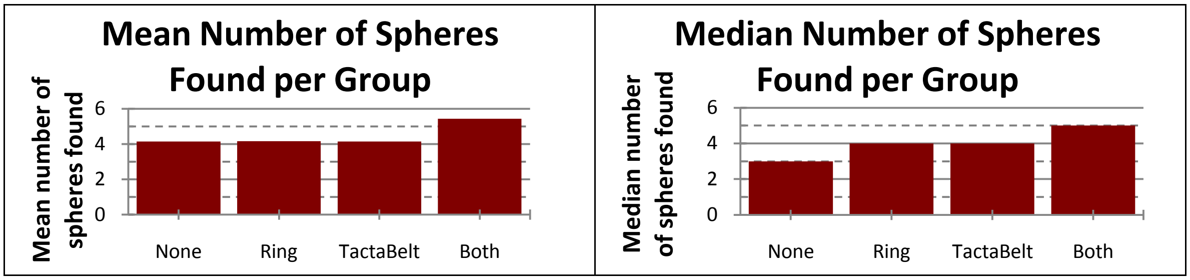


Figure 21: Task time histogram by group.

A single-factor ANOVA comparison for the number of spheres found between groups has also shown no statistically significant difference between groups (F=0.549, $F_{critical}$=3.028, p=0.653). The mean and median for the number of spheres found per group are presented in Figure 22. There appears to be a slight increase in the median value of the number of spheres found as the interface group changes from group "None" (no interface enhancement is used) moving through groups "Ring" and "TactaBelt" (some interface enhancement is used) towards group "Both" (both interface enhancements are used), the latter having the highest median value.



(a) (b)

Figure 22: Mean (a) and median (b) of the number of spheres found per group.

The histogram for the number of spheres found previously presented in Figure 19 is now presented in Figure 23 with subjects from different groups being identified and separated. For all four groups, results seem rather widely spread, with higher peaks which might be a sign of normal distribution.

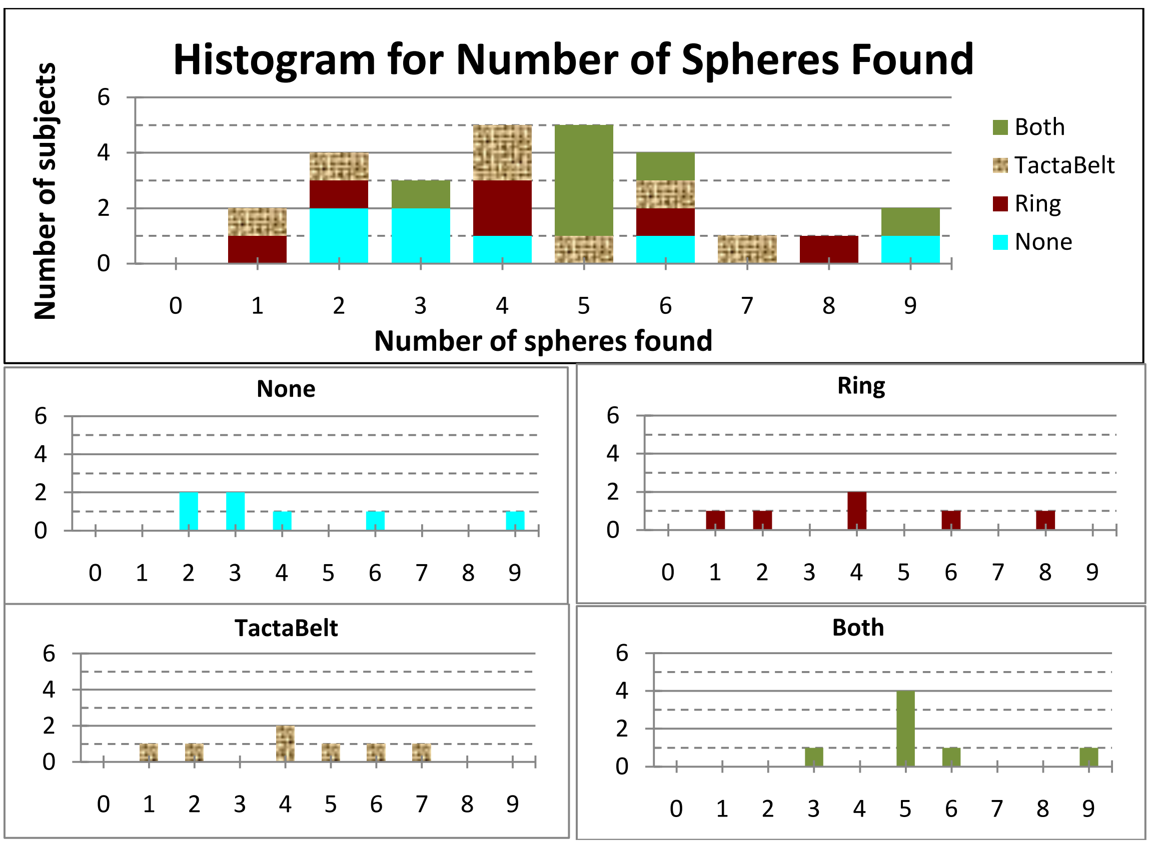


Figure 23: Histogram for the number of spheres found by group.

A single-factor ANOVA comparison for the number of collisions between groups shows a statistically significant difference. The re-application of ANOVA in a pair-wise fashion on the possible combination of groups shows a statistically significant difference between groups “None” and “Ring” (F=5.079, $F_{critical}$=4.844, p=0.025). A statistically significant difference between groups “Ring” and “TactaBelt” was also identified (F=5.079, $F_{critical}$=4.844, p=0.046). No statistically significant difference was found for any of the other pairs of groups. For the group "Both" the cause for this insignificant difference in the results might have been the high variation found in its data. The mean and median for the number of collisions per group are presented in Figure 24 below. Notice the large difference between the results for the ring interface group and each of the other three groups.

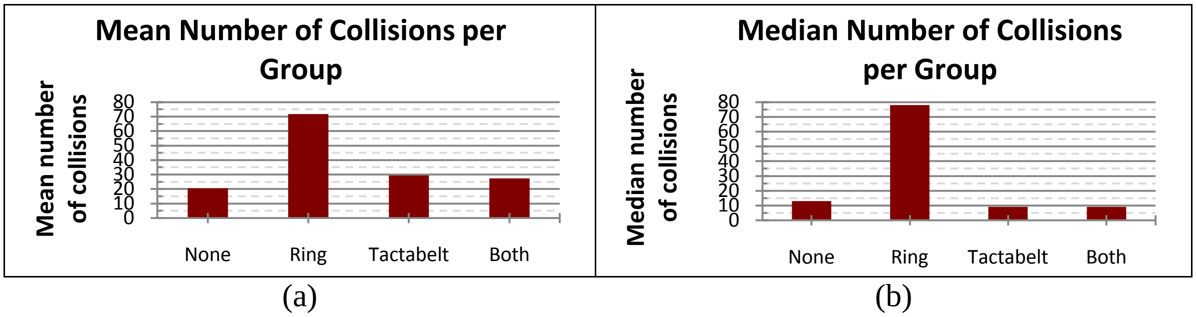


Figure 24: Mean (a) and median (b) of the number of collisions per group.

The histogram for the number of collisions previously presented in Figure 16 is now presented in Figure 25 with subjects from different groups being identified and separated. The results for groups "Ring" and "Both" seem more spread, while the results of groups "None" and "TactaBelt" are more consistent.

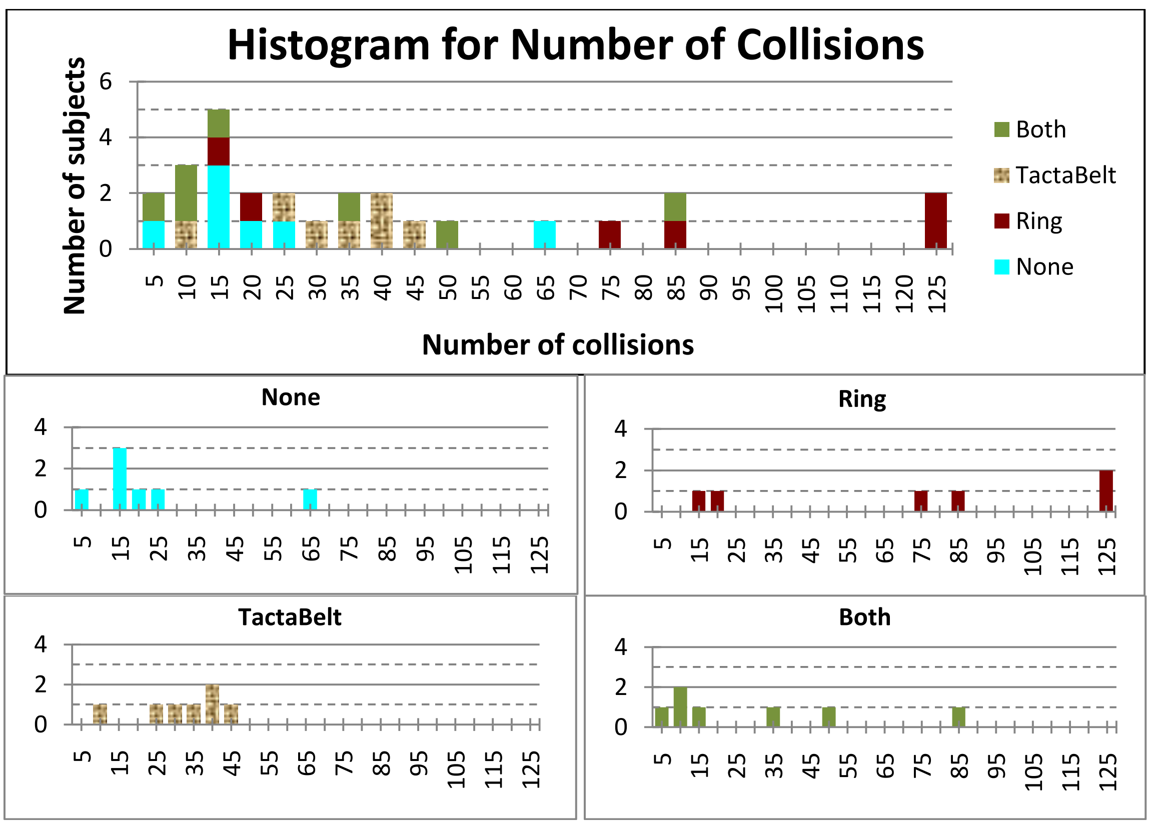


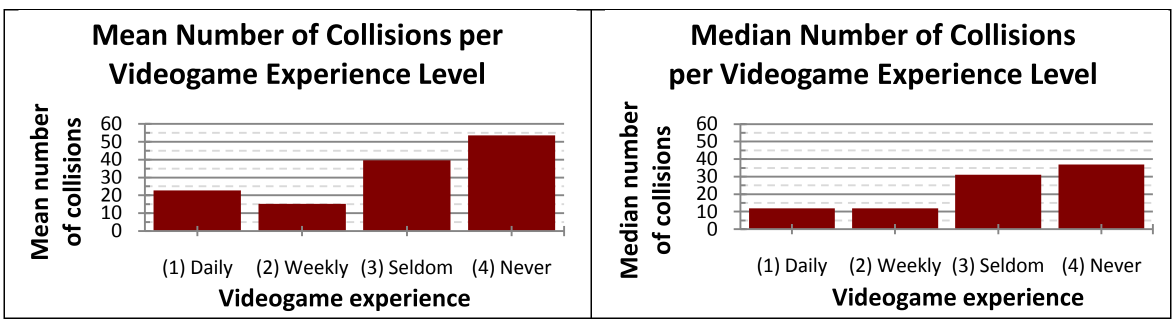
Figure 25: Histogram for the number of collisions with subjects identified per group.

6.5. Analysis of the Relationships between the Main Factors

In order to better understand the causes of variation in performance for different groups, it is important to identify the role that each individual variable plays in the resulting performance values for each group. This section analyzes how much subject background affects performance and how the dependant variables relate to each other. This analysis consists of identifying potential differences between groups caused by nuisance factors such as videogame and robot experience. Additionally, measuring the correlation level between pairs of experimental variables such as task time, number of collisions, and number of spheres found is also presented here in order to support part of our hypotheses.

Nuisance Factors

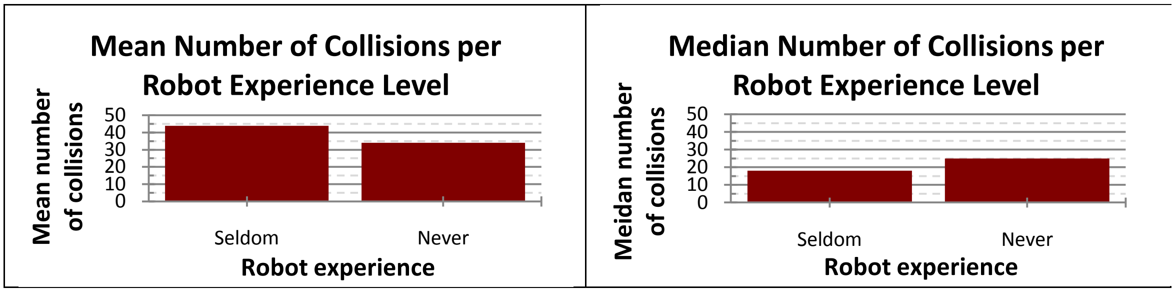
The two nuisance factors considered in this analysis are videogame and robot experience. The first analysis compares videogame experience with number of collisions. The mean and median of number of collisions, separated by levels of videogame experience, are presented in Figure 26. Using a single-factor ANOVA, a statistically significant difference between groups "Weekly" and "Never" was found (F=5.18, $F_{critical}$=4.84, p=0.044). By combining groups "Daily" with "Weekly", groups "Seldom" with "Never" and comparing these two super-groups using a single-factor ANOVA, a statistically significant difference was also found (F=4.38, $F_{critical}$=4.24, p=0.046). The idea was to maximize the amount of subjects in each of the groups, now super-groups, and hence potentially find a statistically significant difference between groups, as shown. This difference indicates, even though neighbor groups were put together, an effect of videogame experience on the number of collisions.



(a) (b)

Figure 26: Mean (a) and median (b) of number of collisions according to level of videogame experience.

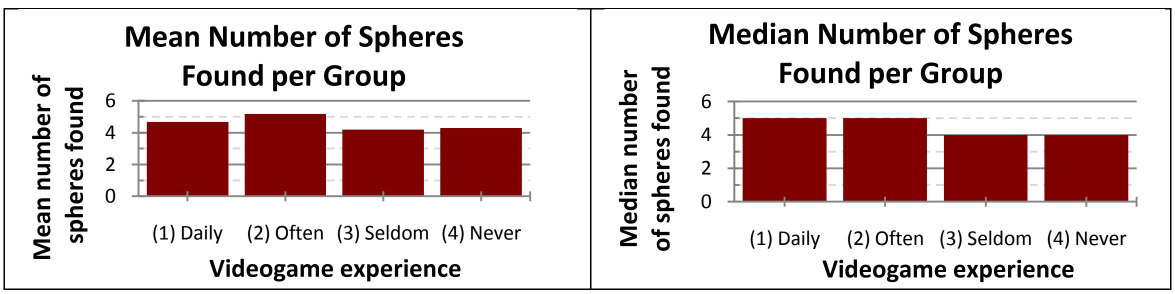
The second analysis, whose results are presented in Figure 27, is between robot experience and number of collisions. No statistically significant difference was found between the group of subjects that had some robot experience ("Seldom") and the group of subjects that had none ("Never").



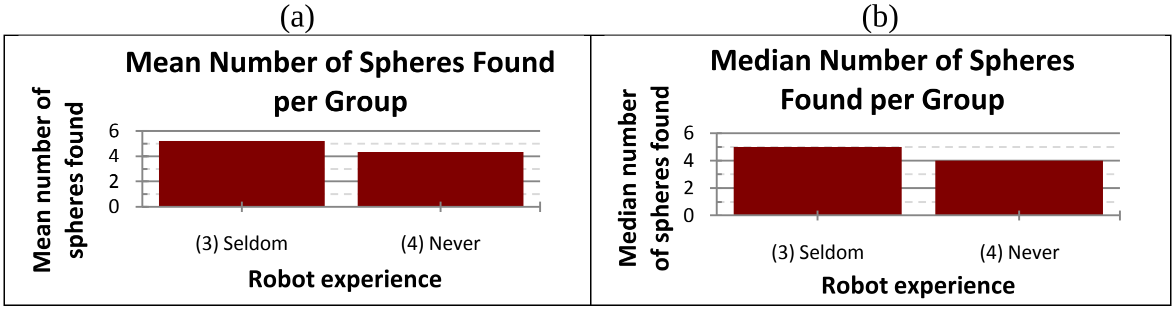
(a) (b)

Figure 27: Mean (a) and median (b) of number of collisions according to level of robot experience.

A similar comparison of the effect of both robot and videogame experience was done for the number of spheres found. For both the robot and the videogame cases, no statistically significant difference was detected on the number of spheres found between groups with different levels experience. According to the medians and means for each group, it seems as though the more experience a subject had, the more spheres he/she would find. These results are presented in Figure 28.



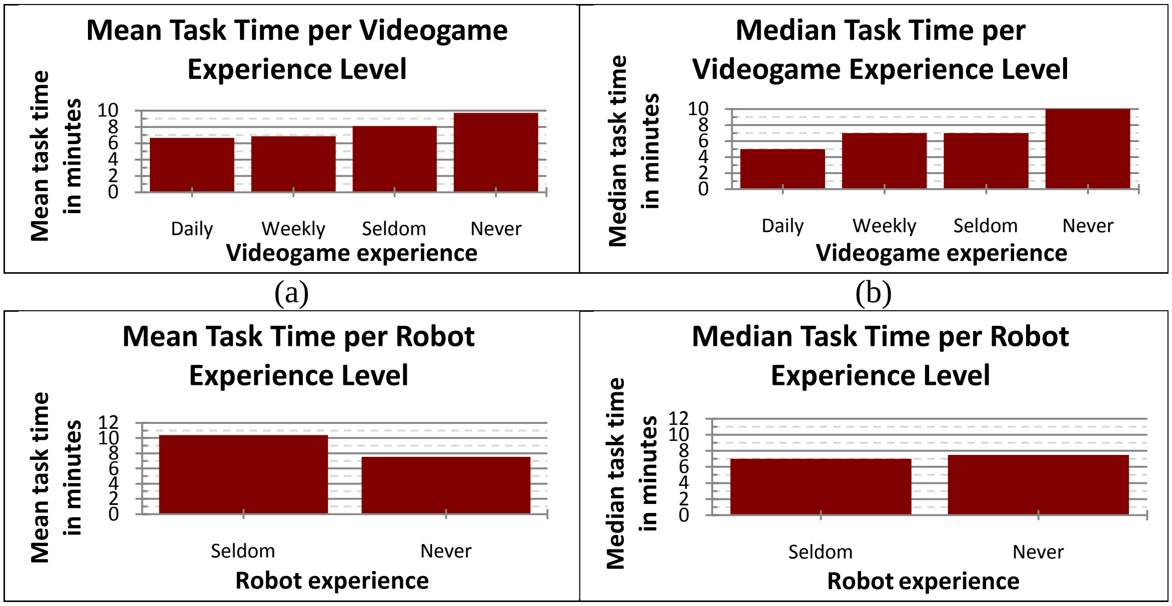
(a) (b)



(c) (d)

Figure 28: Mean (a) and median (b) of number of spheres found according to level of videogame experience; Mean (c) and median (d) of number of spheres found according to level of robot experience.

Videogame and robot experience levels were also tested for having any influence on task time. Once again, no statistically significant difference on task time was found between groups with different robot and videogame experience. From the videogame experience plots of the median and the mean, as shown in Figure 29, it appears that the more experienced the subject was, the faster the task would be performed, but in a very subtle fashion.



(a) (b)

(c) (d)

Figure 29: Mean (a) and median (b) of task time according to level of videogame experience; Mean (c) and median (d) of task time according to level of robot experience.

Correlation Analysis

The results presented here show correlations between pairs of variables that are being considered in our study and its hypotheses. The first one is an expected correlation between task time and number of spheres found by subjects (R = 0.54) as presented in Figure 30.

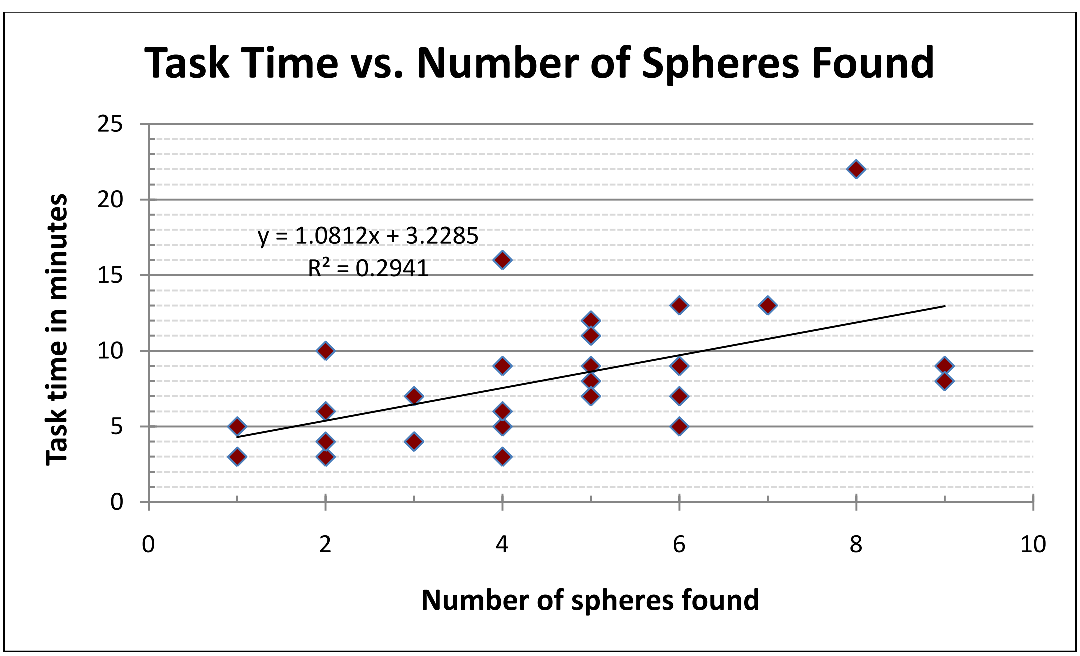


Figure 30: Correlation between task time in minutes and number of spheres found.

The second is a correlation between task time and the number of collisions (R = 0.74). This result was also expected and is presented in Figure 31.

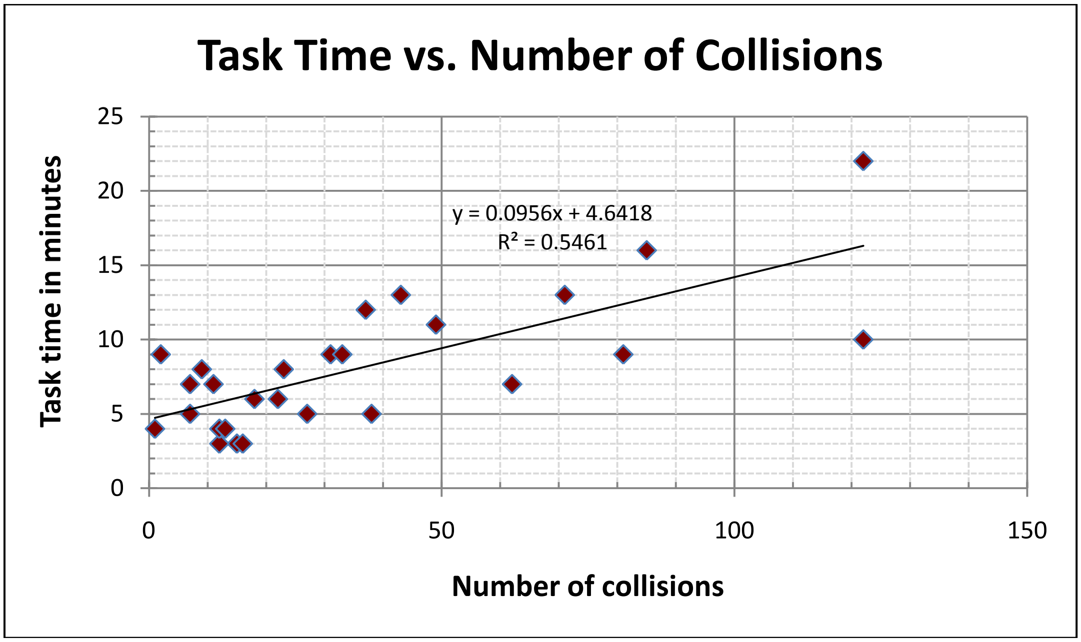


Figure 31: Correlation between task time in minutes and number of collisions.

6.6. Observer notes

Notes were taken by the experimenter while subjects operated the virtual robot through the VEs in the user study. A summary is presented in this section.

- Many subjects returned to the first room of the real environment. The most common reasons for that seemed to be that they either wanted to double check the room after realizing there were more spheres hidden than they thought there would be or because they were simply lost and thought that it was a new room.
- A few subjects reported feeling very annoyed about how the collision with objects would not let them move even if it was not a perpendicular collision. The imprecision of mapping from the virtual to the real world also contributed to this.
- Many subjects went through the environment only once, not going around through all the corners of the main room in search of spheres, but going straight through the exit door. Despite reminders about the exit sign on top of the exit door, subjects would not notice the door was in fact the exit and pass right through it, thus ending their experiment earlier than they intended to. The impression is that these subjects thought the environment was larger than it was and had more rooms than it actually had, perhaps due to their experience with larger videogame environments.
- Most subjects quickly got used to the controls and maneuvered reasonably well, although many of them were deceived by the “stuck” behavior of the collision algorithm, which would not allow the robot to move while colliding, independent of the collision angle. Many videogames allow the player to "slide" when colliding with objects at angles not equal to 90 degrees.

- Two subjects explicitly commented on their doubts about the usefulness of the TactaBelt during the experiment. They felt the TactaBelt was too sensitive or that its range was excessively large. Others would ignore the ring, the belt, and the blue print and only consider them when they could not move the robot around. Despite all the instructions, some would simply not worry about how much the robot was colliding with the environment.
- Less experienced subjects tended to focus operation of the robot on moving the robot around and taking pictures, but keeping the robot camera orientation unchanged most of the time. This behavior often led to a slower search process.
- Only about one fifth of the subjects identified the exit door. About one third to half of the subjects actually cycled around the room more than once.
- Some interesting natural behaviors emerged from subjects during interaction with the HRI system in the experiment. For example, some subjects would turn the chair they were seated in as an attempt to make the robot turn faster. When the robot camera was tilted to its maximum to either left or right and while looking at a corner, a few subjects would turn their heads because they still could not see well what they wanted to.
- Two subjects tried to do more than the robot capabilities and game physics would allow, such as climbing inclined metal panels. Others would not take a picture of a newly found sphere because they thought it to be one they had already seen. Moreover, when lost or for precaution reasons, they would take more than one picture of the same sphere.
- The places where people got stuck the most were in wall corners, door frames, table feet and columns. The two main reasons for that were the imprecision of the virtual world representation of real world objects, and the low visibility of the robot camera, whose field of view would miss thin objects close to the corners of the robot. We feel this, however, is in similar to a real teleoperated robotics scenario.
- During the user study, the application crashed once, but before the training session had started.
- One collision bug was found by a couple of users: in a specific point of the map the robot camera would pass through an inclined surface, but not the robot itself.

6.7. Sketchmaps

The results for sketchmaps are presented according to the different grading factors explained in the User Study section. Grading was done by the experimenters. The results for map goodness are presented in Figure 32. Almost half of the subjects failed to make a good representation of the environment, and had their maps graded as 1 or 2. The mean map goodness was of 2.78 with standard deviation of 1.9 and median of 3.

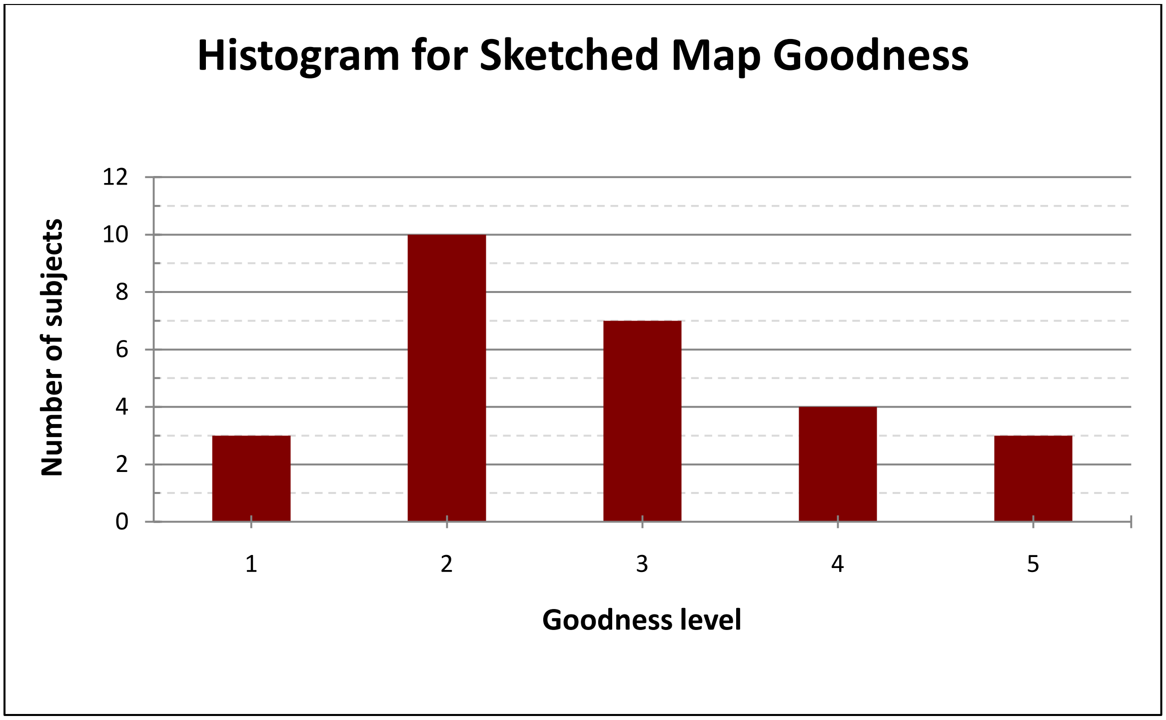
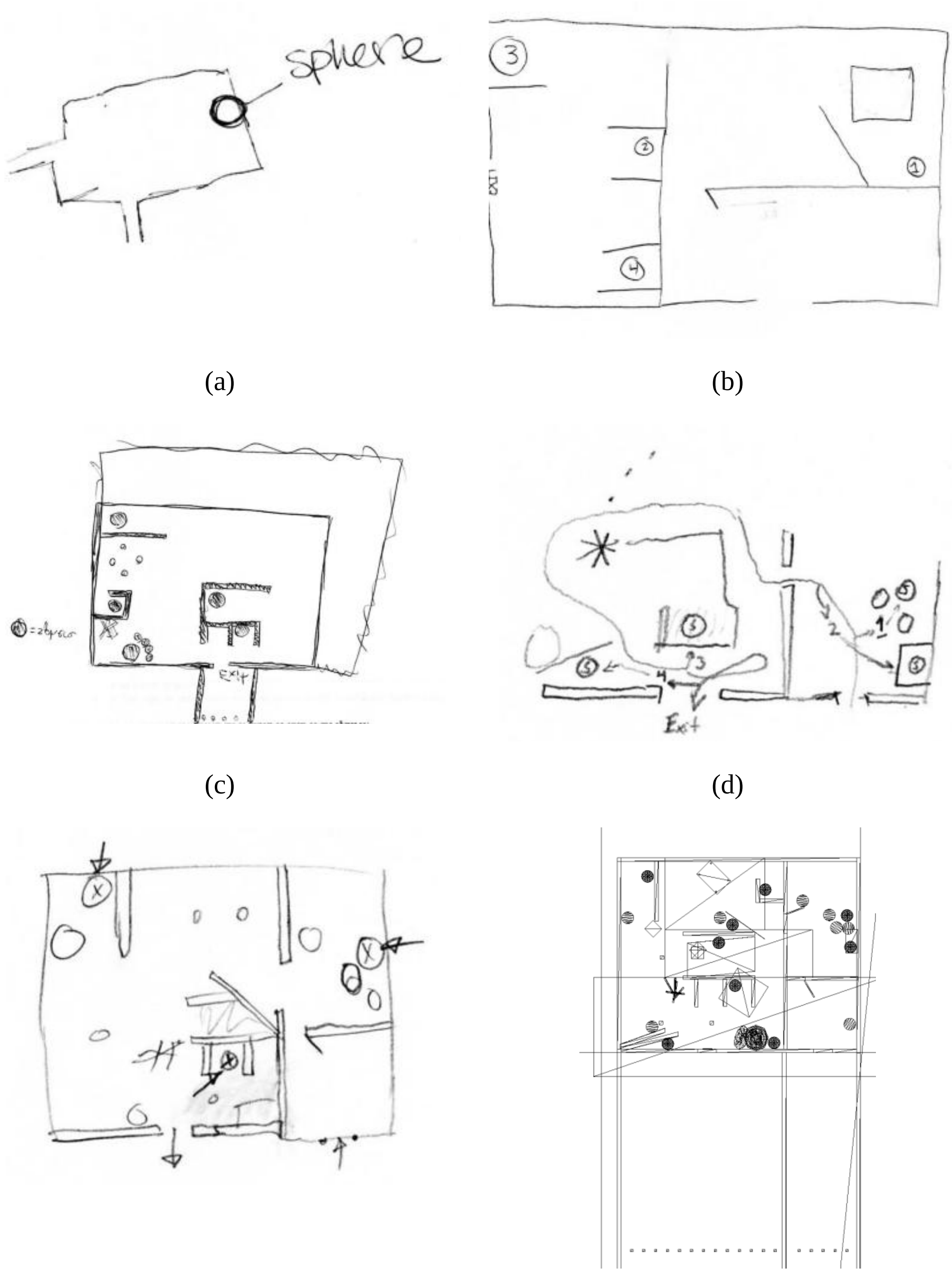


Figure 32: Histogram for map goodness.

Maps graded with 1 provided no help as a guidance tool through the environment. Maps graded with 2 had the description of a few features of the environment but represented it with a large number of mistakes in terms of spatial representation. Maps graded with 3 had some features of the environment well placed and described in text, but still had major errors in their sketches, such as the wrong number of rooms and doorways. Maps graded as 4 were describing the environment correctly except for the misplacement of some objects and walls. Maps graded as 5 had sketched the environment almost completely correct and all the objects found were in the right place. Samples of maps sketched during the experiment and graded differently are presented in Figure 33.

Some subjects also added extra features to their descriptions of the scene, by drawing the approximate path they went through during the search task (Figure 33.d) or the order with which they found the spheres and how they relate to the pictures taken by the robot camera (Figure 33.b and c).



(a) (b)

(c) (d)

(e) (f)

Figure 33: Sketchmap samples for different grades: (a) goodness grade = 1; (b) goodness grade = 2; (c) goodness grade = 3; (d) goodness grade = 4; (e) goodness grade = 5; (f) original map.

The environment was considered as having 24 different walls. The graph results for the number of walls sketched shows an average of 16 walls sketched per person, as shown in the histogram of Figure 34. For poor sketches, the number of walls sketched varied to more or less than the mean. More walls being drawn than the real existent number indicate that more rooms had been represented than there actually were in the environment. The mean number of walls reported was of 15 with standard deviation of 5 and median of 16.

When comparing groups of subjects whose maps where given different goodness levels, and using a single-factor ANOVA, a statistically significant difference in the number of walls was found between groups with map goodness levels of 1 and 3 (F=6.73, $F_{critical}$=5.32, p=0.032) and between groups with map goodness levels of 1 and 5 (F=21.16, $F_{critical}$=7.71, p=0.01).

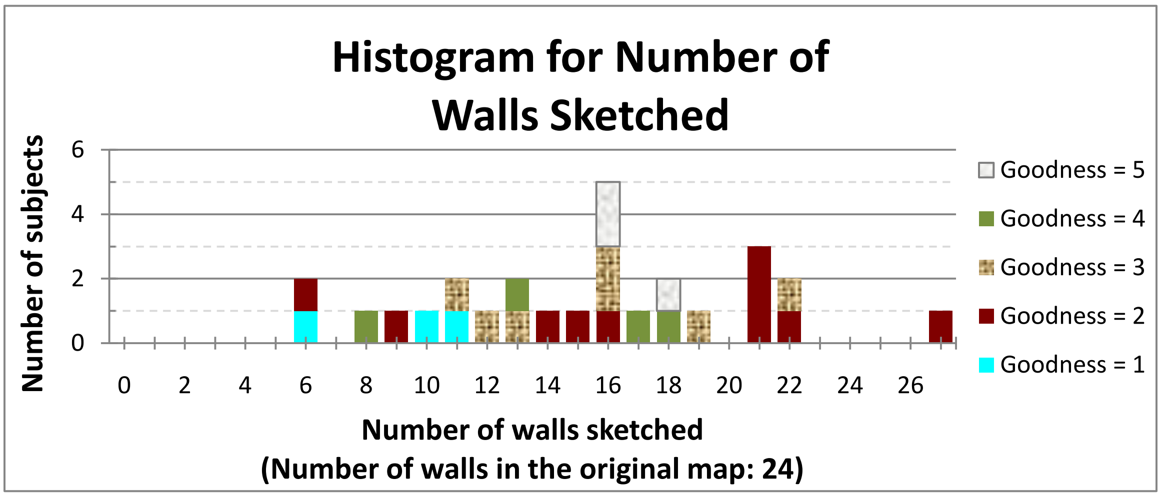


Figure 34: Histogram for number of walls sketched by subjects.

A similar result was obtained for the number of doorways sketched. However, in this case the number reported was more accurate. About half of the subjects correctly reported the four doorways present in the original environment. Figure 35 presents the results. Notice that there was also overestimation of doorways once again due to over complication of the environment spatial structure by some subjects. The mean number reported was 4 with standard deviation of 1 and median of 4.

Having map goodness level as the grouping criterion to compare the number of doorways sketched, a statistically significant difference was found when applying a single-factor ANOVA between groups with map goodness levels of 1 and 4 (F=7.10, $F_{critical}$=6.61, p=0.045) and between groups with map goodness levels of 1 and 5 (F=16, $F_{critical}$=7.71, p=0.016). This indicates that doorway sketching is a dominant component in the evaluation of map goodness.

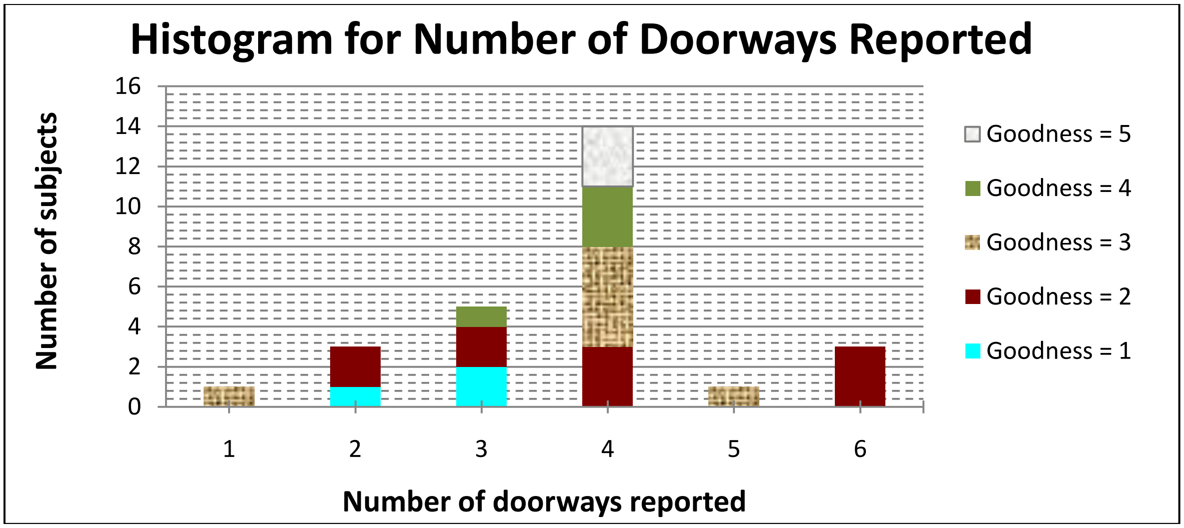


Figure 35: Histogram for the number of doorways reported on sketchmaps.

The total amount of debris in the environment was identified as 17 distinguishable pieces: eight oil barrels, four sets of metal/wooden panels, one set of tires, one set of wood sticks, one table and two columns. The amount of debris reported varied reasonably. Additionally, some people did not report any pieces of debris at all. Figure 36 presents the amount of debris reported by subjects. The mean number of pieces of debris reported was 6 with standard deviation of 4 and median of 5.

Still considering map goodness level as a grouping criterion, a statistically significant difference in the number of debris items sketched was found when applying a single-factor ANOVA between groups with map goodness levels of 1 and 4 (F=58.56, $F_{critical}$=6.61, p=0.0006) and between groups with map goodness levels of 1 and 5 (F=8.08, $F_{critical}$=7.71, p=0.047).

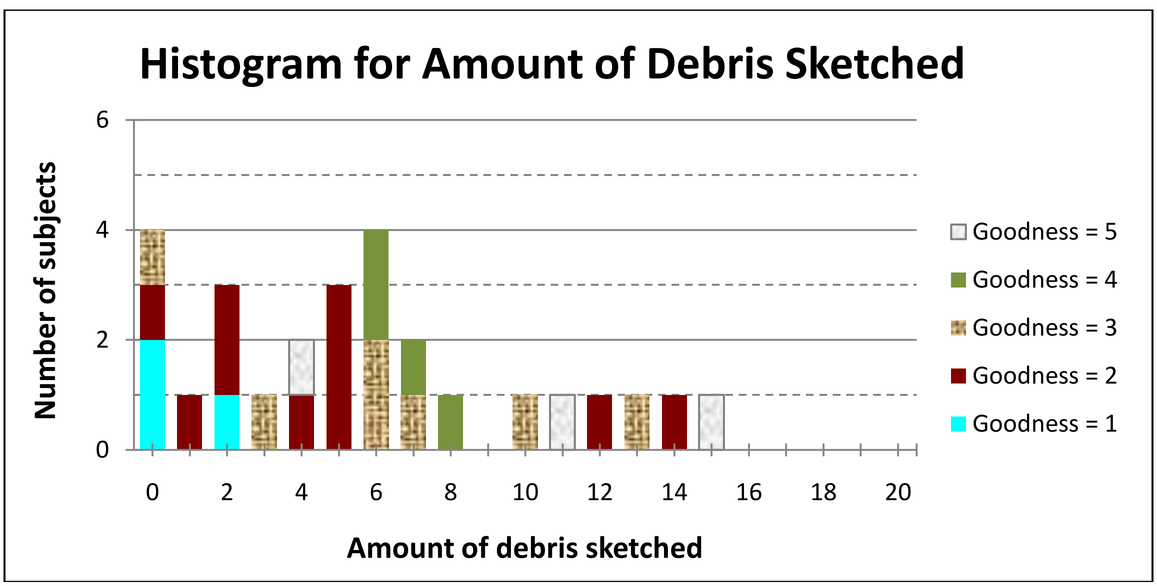


Figure 36: Histogram for amount of debris reported by subjects.

Last, the relative position of objects results are presented in Figure 37. Objects with a larger count for inter-object spatial relations tended to have received a higher goodness score.

Additionally, statistically significant differences in the number of debris sketched were found when applying a single-factor ANOVA between the following map-goodness level groups (p<0.05):

- 1 and 3 (F=6.76, $F_{critical}$=5.32, p=0.032);
- 1 and 4 (F=137.28, $F_{critical}$=6.61, p=7.9×10^{-5});
- 1 and 5 (F=28.12, $F_{critical}$=7.71, p=0.0061);
- 2 and 4 (F=17.77, $F_{critical}$=4.84, p=0.0014);
- 2 and 5 (F=10.13, $F_{critical}$=4.96, p=0.0098);
- 3 and 4 (F=12.32, $F_{critical}$=5.12, p=0.0067) ;
- 3 and 5 (F=6.37, $F_{critical}$=5.32, p=0.036).
-

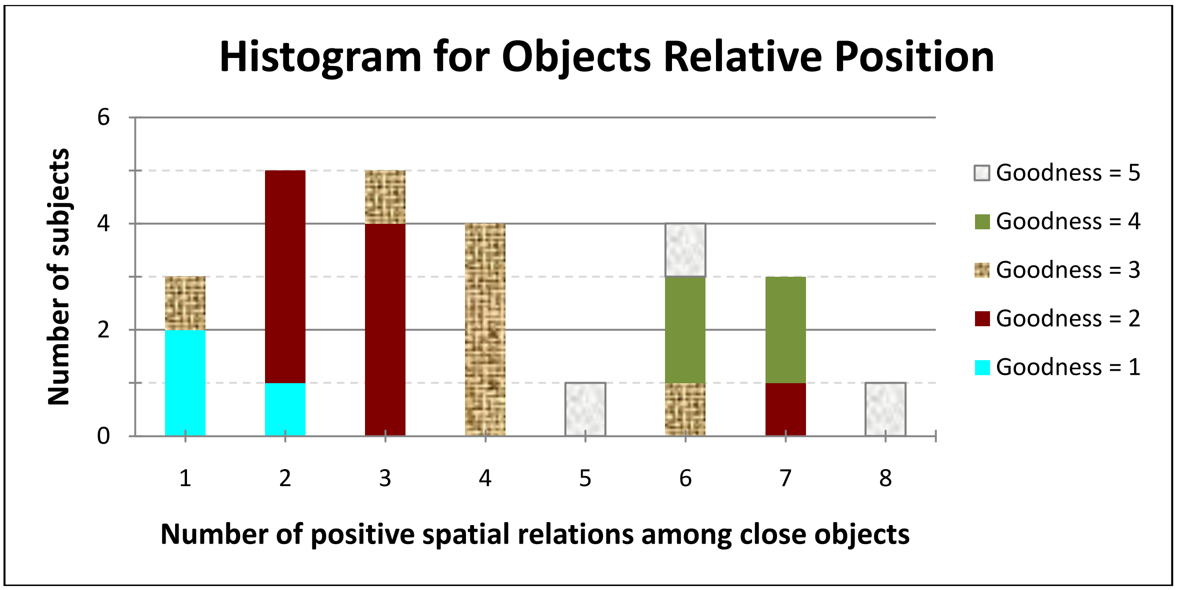
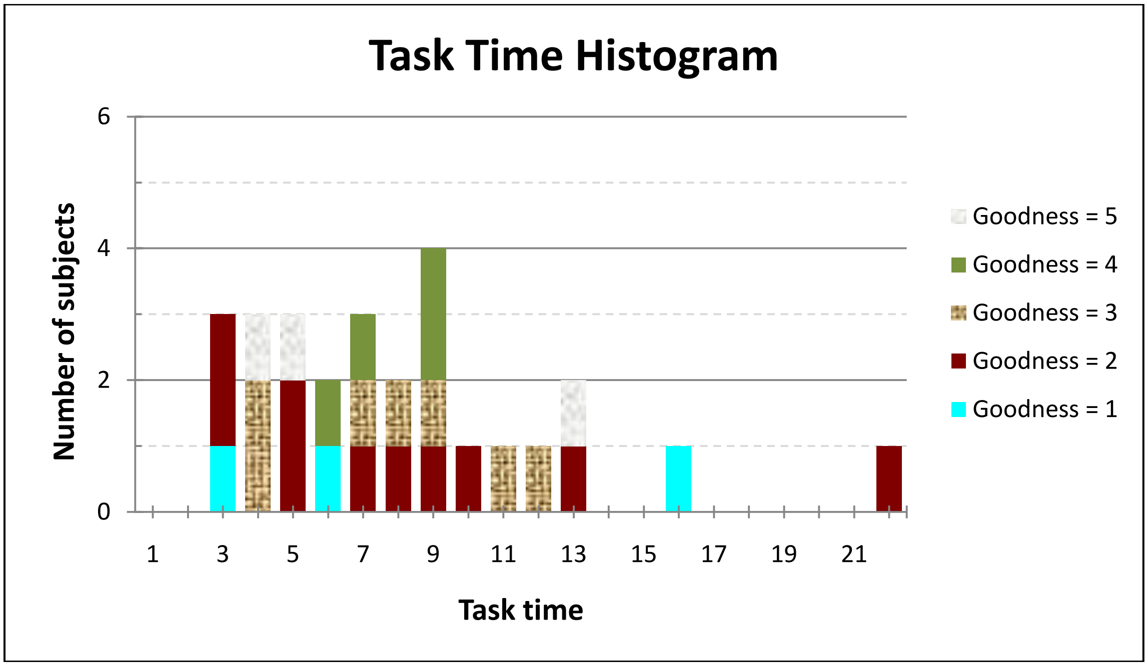
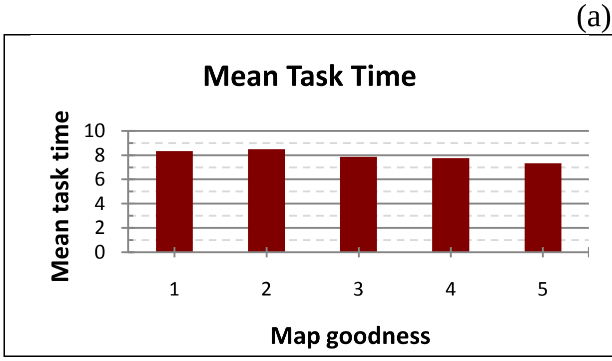


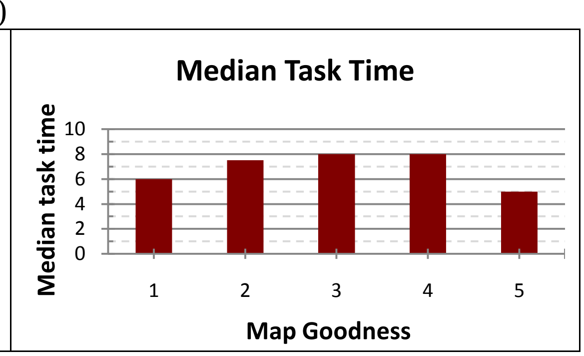
Figure 37: Histogram for the relative position of objects.

Map goodness was also compared to task time, videogame experience, robot experience and interface groups. When comparing it to task time, no statistically significant difference amongst groups was detected by a pair-wise single-factor ANOVA. Good and poor maps were sketched by subjects who spent from 4 minutes to 20 minutes in the environment. Figure 38 presents the task time histogram colored according to map goodness. It also presents the means and medians of map task time according to the groups of subjects whose maps were graded to a certain goodness level.



(a)





(b) (c)

Figure 38: Map goodness vs. task time: (a) task time histogram colored according to goodness of subjects' sketchmaps; mean (b) and median (c) task time for subjects rated at different map goodness levels.

When comparing map goodness to videogame experience, a slight shift in the concentration of subjects at certain experience levels is present as the goodness level increase from 1 to 4. These results are presented in Figure 39. However, for a goodness level of 5, such a pattern is not perceived. Again, no statistically significant difference was found amongst groups by a single-factor ANOVA applied pair-wise. These results seem to indicate that videogame experience might have an impact on the quality of the subject sketchmaps.

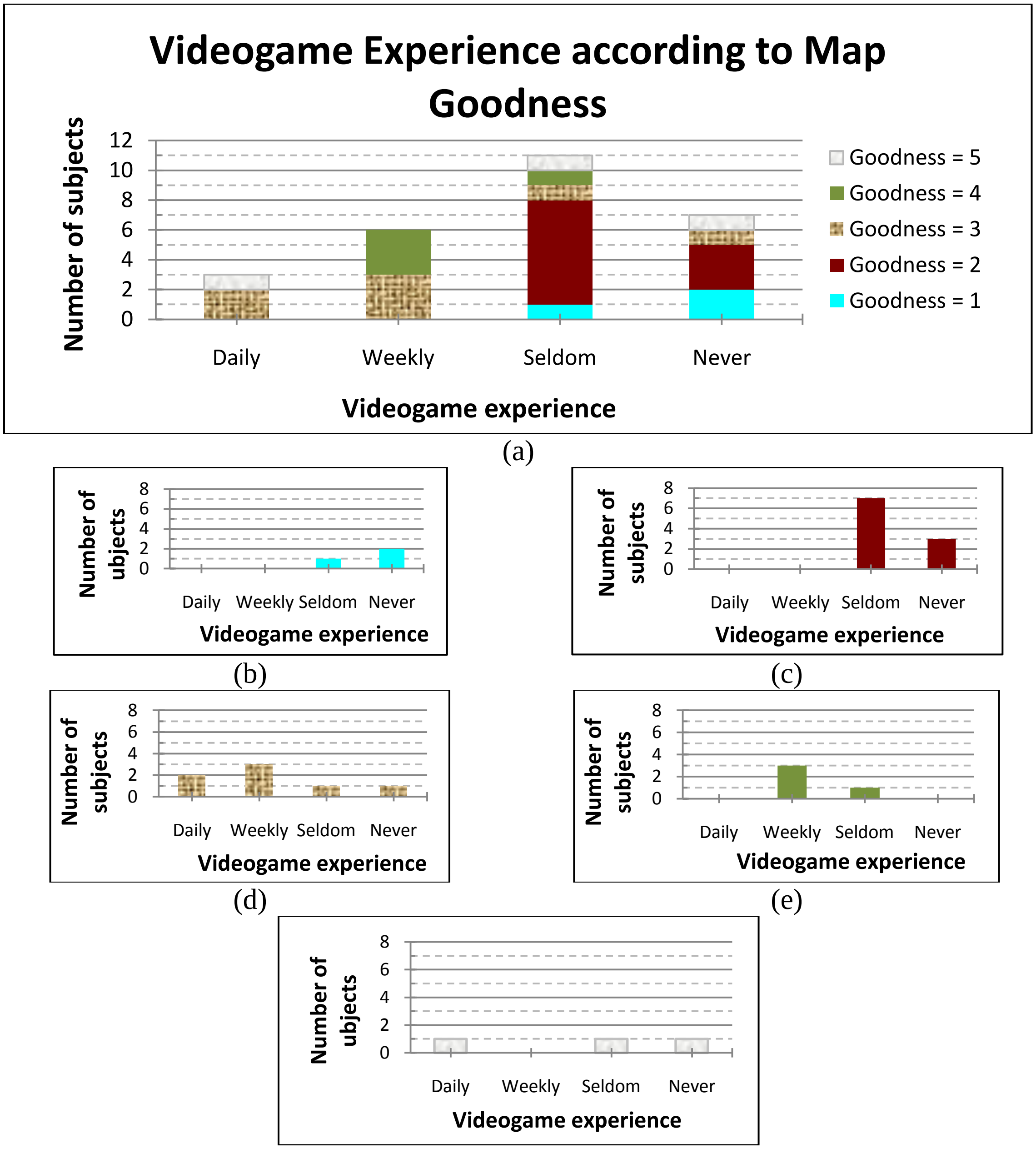


Figure 39: Distribution amongst different levels of videogame experience is presented for all goodness levels at once (a) and individually for map goodness levels 1 (b), 2 (c), 3 (d), 4 (e) and 5 (f).

The mean and median for map goodness according to videogame experience are presented in Figure 40 below. For the mean, a small growth in map goodness as videogame experience increases is perceptible. When applying a single-factor ANOVA to the four groups with different levels of videogame experience, no statistically significant difference was found.

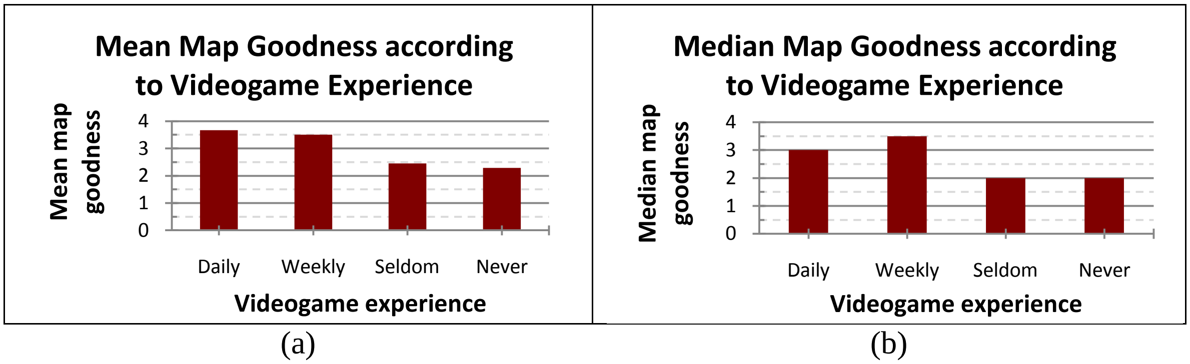


Figure 40: Mean (a) and median (b) of map goodness levels according to videogame experience.

A similar analysis was done to compare map goodness to robot experience as shown in the histogram of Figure 42. Even though there is an increase in the mean and median of the map goodness as robot experience increases, as presented in Figure 41, no statistically significant difference was found between the two available groups.

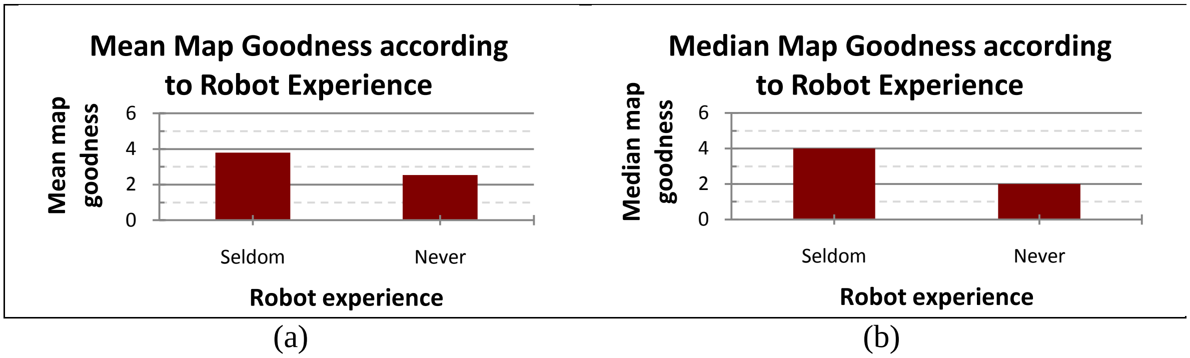


Figure 41: Mean (a) and median (b) of map goodness levels according to robot experience.

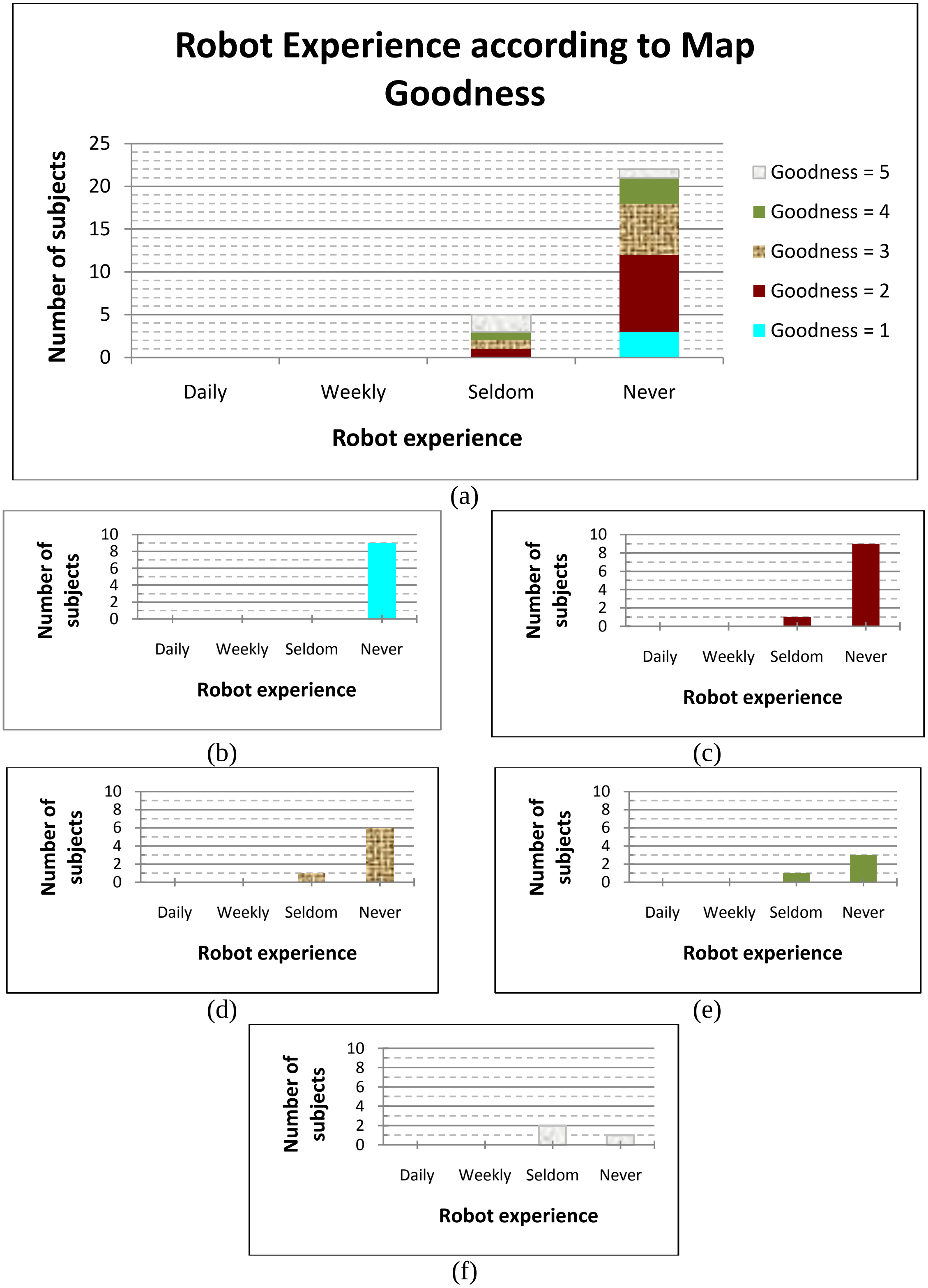


Figure 42: Distribution amongst different levels of robot experience is presented for all goodness levels at once (a) and individually for map goodness levels 1 (b), 2 (c), 3 (d), 4 (e) and 5 (f).

Last, a comparison between map goodness and the type of interface used was done. From Figure 43, a large variation is noticeable in the quality of maps for group “Ring”. Groups “None” and “TactaBelt” had similar map goodness distribution levels. Group “Both” had a map goodness distribution more highly centered.

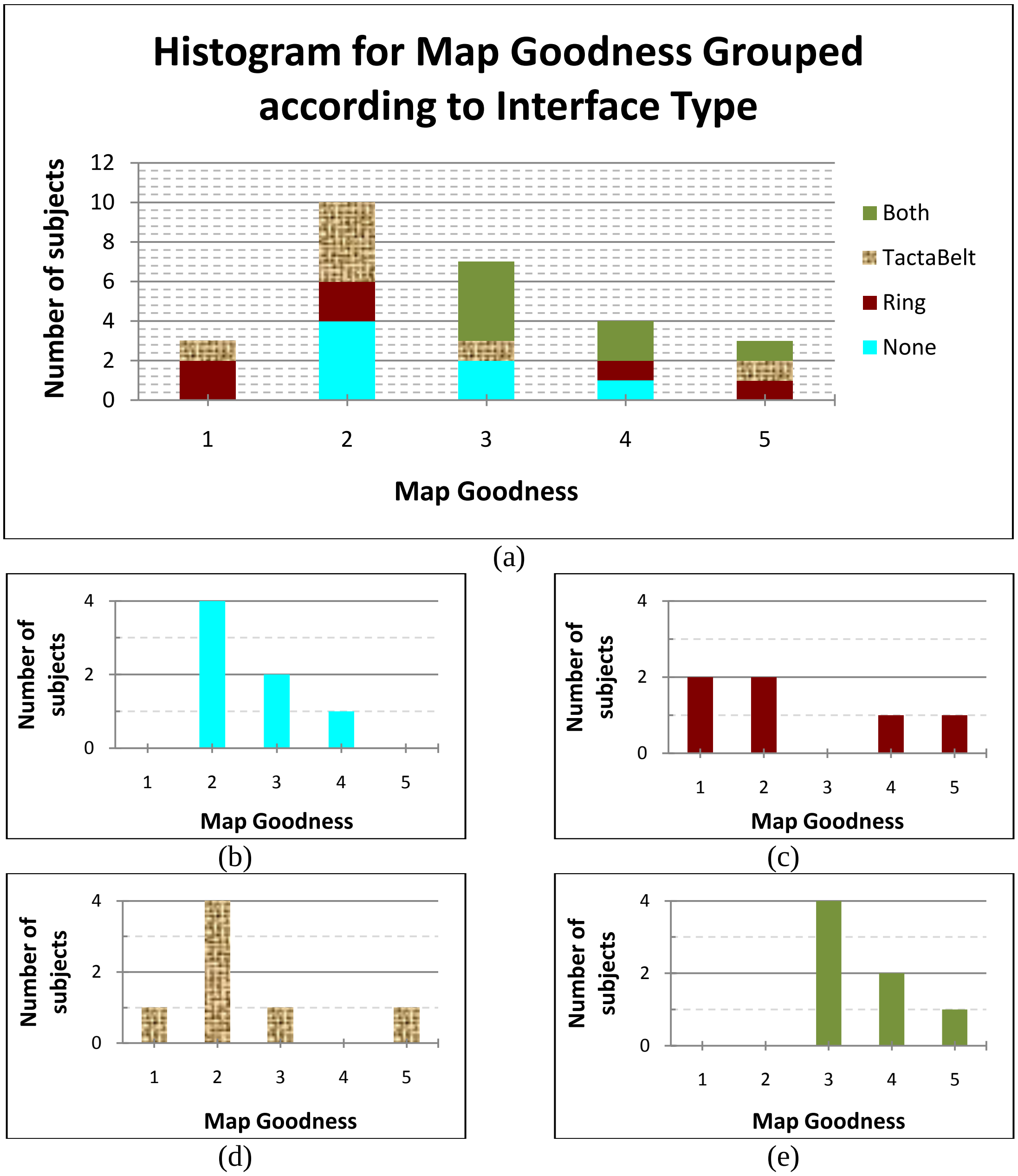


Figure 43: Distribution of subjects over different map goodness levels according to the type of feedback interface is presented for all interface feedback groups together (a) and separately for groups “None” (b), “Ring”(c), “TactaBelt” (d) and “Both” (e).

The mean and median of the map goodness levels for the subjects grouped according to interface makes the abovementioned difference more evident in distribution result for group “Both”. An ANOVA performed pair-wise for each combination of two groups indicated a statistically significant difference only between groups "None" and "Both" (F=5.65, $F_{critical}$=4.65, p=0.035, p<0.05). An inverted-axis plot of Figure 42 is presented in Figure 44 more clearly represents this variation for group "Both". Notice in the group "Both" graph column the absence of sketchmaps rated with goodness levels 1 or 2. This is an important result, because it indicates the positive effect caused by the CPF interfaces on subject's situation awareness levels.

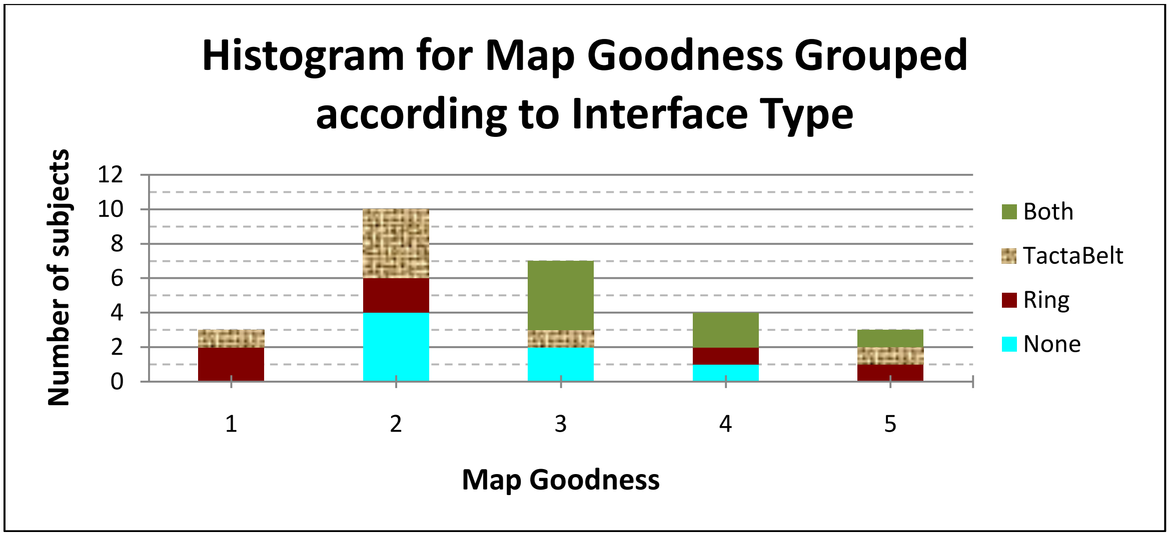


Figure 44: Histogram for interface types colored according to levels of map goodness.

6.8. Subjects Comments

Most subjects left comments about the experiment when asked in the post-questionnaire. They are summarized in this section.

Ten of the subjects found the robot slow. Four of them found that the TactaBelt was vibrating too much and making the pilot task more confusing than helpful. There were two comments asking for reduction in the sensitivity of the belt.

Four comments praised the helpfulness of the blue lines on the ground. Four other comments complained about their imprecision. Another four comments mentioned the difficulty in understanding when the robot was able to pass through certain parts of the environment. One comment mentioned the environment was excessively tight in terms of collisions. Another mentioned it was hard to turn left or right while moving the robot forward.

There were also some complaints about the camera view. Four comments were written about how hard it was to see the camera when it was tilted to the side. Two other comments mentioned that the camera would not match the sensor feedback and that it would not show the robot front, making it hard to estimate collisions through the video feed. One subject asked for a switch button to enable inversion of the pitch movement of the robot camera.

One subject admitted not having moved the camera much, making the search for the spheres more difficult. Another suggested the information of the amount of spheres hidden in the environment before the experiment starts. Two subjects thought more videogame experience would have helped them perform better.

7. Discussion

Comments on the results and how they relate to the hypotheses initially claimed are presented in this section.

7.1. Population information

The population results seem to indicate a fair distribution of gender over age. However, male subjects seemed to have more experience with robots and videogames than females.

7.2. Homogeneity

Even though at least two subjects from both genders were present in all four groups, some groups had twice as many subjects from one gender than from the other. If there is any significant difference in performance caused by gender, groups may have been affected by it, especially groups “None” and “Both”.

All groups had approximately the same average age, when not considering outliers, except group “Both” which had a higher age mean. This fact, allied with the fact that most of the subjects in group “Both” were male, and that males have generally more experience with robots and videogames, may indicate a bias in the performance results for this group. The set of subjects in this group might have been able to perform better than the subjects in other groups even if all groups were to use the same CDF interface. This possibility is partly confirmed by the high experience mean of this group as presented in Figure 13.

It is interesting to notice that group “Ring” has the lowest mean in videogame experience and the highest in experience with robots. These features may indicate that while other groups are being more affected by any benefits of being an experienced gamer, this group is not. This may be one of the reasons why this group had poor results compared to other groups in the number of collisions and task time.

Additionally, only two of the groups had members with experience using robots. Hence, if robot experience had any effect on subject performance, this effect may have biased the results between this pair of groups and the other two groups. However, our results seem to only indicate a slight statistically insignificant positive effect of videogame experience, not robot experience, on user’s performance in terms of number of collisions, number of spheres found and task time.

7.3. General results

The general results for task time, as presented in Figure 15, indicate that the distribution of subjects may be following a multimodal behavior. This multimodality appears to be associated with the number of cycles subjects have gone around the environment in search for the red spheres.

The first set of subjects would be identified with a task time of approximately four minutes and would have gone through the environment only once and without checking all the corners before passing through the exit door. The eleven subjects that spent 3 to 6 minutes in the task environment had a mean of 2.9 spheres found.

The second set would be identified with a task time of around nine minutes and would represent subjects that had gone through all the corners before exiting the environment, but did not explore all the perspectives for each position. The ten subjects that spent 7 to 10 minutes in the task environment had a mean of 5.4 spheres found, almost one sphere unit higher than the global mean value of 4.48.

The third set would comprise a few subjects who spent around 13 minutes and performed and in-depth search of the environment, often finding most of the spheres. The four subjects who spent 11 to 13 minutes in the task environment had a mean of 6.5 spheres found, 2 sphere units higher than the global mean value of 4.48.

The fourth and last set of subjects would be composed of outliers that had encountered severe problems during their task such as map disorientation or excessive collision with the environment. The high number of collisions was confirmed by a data check for the two outliers located at 16 and 22 minutes. Observer notes confirmed map disorientation and total lack of experience from the subjects located in this group. On the other hand, the mean number of spheres found for these two subjects was of 6.

A similar group distinction may also be seen in Figure 16, where the four peaks for the above mentioned groups are again identifiable. Some multimodal behavior is also perceptible in Figure 18 and Figure 19, although the peaks should appear in an inverse order for these graphs, and they look less accentuated.

The training session time had some variation depending on the subject being run. Such time flexibility had to be present so that subjects would all start the real experiment with approximately the same level of understanding of the interface.

7.4. Performance Comparison between Groups

When comparing the task time results amongst groups, the results seem to indicate that the use of CPF interfaces lead to an increase in task time, as seen in Figure 20, although no statistically significant difference among group results was encountered. The individual plot of each group task performance in Figure 21 shows some normal behavior for some groups, but a larger pool of subjects would be necessary to achieve more conclusive results.

Similarly, when comparing the number of spheres found group-wise, the use of CPF interfaces seemed to have caused a slight increase in the number of spheres found (see Figure 22). Nevertheless, once again, no statistically significant difference among group results was found. When looking at Figure 23, it is noticeable that the distribution peaks at 2-3, 4, 4 and 5 for the number of spheres found relates to the data coming from groups “None”, “Ring”, “TactaBelt” and “Both” respectively. With enough data, this may indicate that the use of CPF interfaces might help increase search performance and situation awareness and indirectly imply in a reduction in subjects’ cognitive load caused by robot navigation. Additionally, since group “Both” had the largest mean as shown in Figure 22, this would also indicate that the combined use of both the feedback ring and the TactaBelt interfaces had the best performance in terms of spheres found and hence and increase in situation awareness and a reduction of the interface cognitive load on the operator. Nevertheless, the fact that group “Both” was composed of more experienced subjects might have biased the results and might deny us from reaching such a conclusion with complete certainty.

The comparison of the number of collisions between groups led to some interesting results as presented in Figure 24 and Figure 25. It seems that the groups using the ring interface had more collisions than the groups not using it. This would be one possible explanation for the fact that a statistically significant difference on the number of collisions was found when comparing groups not using the ring - groups “None” and “TactaBelt” - with group “Ring”, but no statistically significant difference was found when comparing “Ring” to the other group that was also using the ring, which was group “Both”.

This decrease in performance caused by the ring interface may be due to the fact that the ring graphical representation covers the surrounding area of the robot avatar, making it more difficult for subjects to see the map, represented as blue lines on the ground. However, the excessive difference between groups “Ring” and “TactaBelt” in number of collisions may have been caused by the fact that group “Ring” was composed of mostly less videogame-experienced subjects and more robot-experienced subjects. Although robot experience appeared to have had little or no impact in subjects’ performance in terms of the number of collisions, videogame experience did have an effect, as shown in Figure 26.

When analyzing the collision data per interface group, statistically significant differences between groups “Ring” and “TactaBelt” and between groups “None” and “Ring” were found, but not between any other pairs of groups. These results seem to lead to the conclusion that something in the ring interface needs to be improved in order to actually make it beneficial for the operator. With regard to the TactaBelt, the results only allow us to reject the hypothesis that the TactaBelt significantly improves performance with regard to the number of collisions. But the TactaBelt has improved user collision-avoidance performance when used together with the ring interface. However, compared to the results of only using the ring interface no performance improvement was detected. Similarly, no statistically significant results were found when the TactaBelt was used without the Ring interface and compared with the control ("None") group. Assuming the subject composition of group “Ring” had not significantly affected its performance scores, it is viable to claim that the used of TactaBelt negatively affects less subjects’ performance than using the graphical ring in terms of number of collisions. Additionally, we may claim that the graphical ring alone, as it was presented here, was not a viable solution for a CDF interface.

7.5. Nuisance Factors and Correlation Analysis

Neither robot experience nor videogame experience seemed to have any major effect on the results obtained on the number of spheres found, although the mean plots tend to slightly increase along with experience in both cases.

In terms of the effect on task time, once again neither videogame experience nor robot experience had a statistically significant difference among groups. Nevertheless, videogame experience results for mean and median have shown a slight decrease in task time as videogame experience increased.

The correlation results point to the conclusion that an increase in task time is more associated with an increase in the number of collisions than to an improvement in the number of spheres found, due to the stronger correlation between task time and number of collisions found as presented in Figure 31. This means that a large increase in task time is more likely to be caused by an inexperienced operator who collides frequently with the environment than because of a more-experienced operator doing an in-depth search around the environment. This validates our claim that, by reducing the number of collisions by using new feedback interfaces, subject performance should be enhanced.

The correlation between task time and number of spheres may only point out that more time spent on the search task equates to more spheres being found by subjects, even if accidentally.

7.6. Sketchmaps

The map results show that half the subjects did not know how to recall and draw the environment they were in.

The results for relative position of objects show that this was the main factor used in grading maps by the map evaluator amongst the factors considered in our map analysis in section 6.7. Other factors, such as the number of walls, doorways, and the amount of debris seem also to have some relation with map goodness.

The amount of time spent on the task and the quality of the map produced seem to be unrelated according to our results. The same can be concluded for robot experience. Videogame experience, however, seems to have some effect on the quality of the sketchmaps.

The most interesting results obtained from sketchmaps was the increase in the quality of the maps due to the combined use of the TactaBelt and Ring feedback interfaces. Interestingly, separately they do not show a similar increase in map goodness to when they were used together as seen on Figure 44. This may be indicative that this interface may actually improve operator situation awareness, or it may be a consequence of group “Both” being a slightly more experienced group.

8. Conclusion

The improvement in number of spheres found, although not statistically significant, indicates the potential that the proposed interface has on reducing the cognitive load on operators. Additionally, the interface benefits may impact task time, leading to slower, though more effective, performance. But this implication was not validated by statistical results. These variations may be the result of the heterogeneous distribution of subjects among groups, which caused improvements in the results obtained for group “Both”.

In terms of collisions, it appears that the current version of the ring feedback interface needs to be improved. The results are opposite to what our first hypothesis stated. A more in-depth study must be performed in order to understand whether such a significant decrease in performance was due to the interface itself or to the fact that group “Ring” was mostly composed of inexperienced videogame players.

The subject comments indicate that a better fine tuning of the sensing radius and vibratory levels for the TactaBelt should be established according to the environment dimensions. The levels used during this study appeared to be appropriate in the training room, but generated a large number of complaints in the task environment. This fine-tuning will be explored in future studies.

It is important to highlight the relevance that the homogeneity check amongst subjects had on the results of this study. Most research results do not tend to perform, or at least do not report, such verification. If our current results had not been verified for bias, they would probably have led us to different conclusions. The soundness of the random process used in a research study such as this one is of the utmost importance to guarantee the validity of its conclusions. This type of more refined analysis of the population is a topic that should be given more attention in general by the research community.

The fact that group “Both” had drawn better maps than all other groups and that the TactaBelt had no negative impact for all experiments run indicate that the use of this interface in conjunction with other graphical CPF interfaces may improve operator’s situation awareness without detriment of cognitive load.

We believe that vibro-tactile feedback interfaces have the potential to improve HRI interfaces, which today are mainly limited to graphical output. With some enhancements to the application interface and being run during a longer period of time with a larger number of subjects, a future continuation of this study would likely be able to reduce the amount of population-related bias and increase the confidence level in the obtained results. A within-subjects study would also help alleviate this threat to statistical validity, but would then require us to generate and validate several test environments, so that learning effects do not compromise its validity.

Other areas of future work include the addition of more feedback mechanisms from the robot to the operator that are already commonly used, such as CO_2 level meters, multi-camera views, and flashlights. Additionally, we noted that a better control and review of the experimental process could be achieved by the use of instructional videos and by recording of subjects during the task. This is planned to be introduced in future user studies to reduce the chances of human error by the experimenter, and to allow more detailed annotations per subject.

9. Acknowledgements

Thanks to my advisor, Dr. Robert Lindeman and his infinite patience and dedication and to Dr. Matthew Ward for his experience sharing and guidance during the entire research process.

Thanks to the Computer Science Department of the Worcester Polytechnic Institute that has been funding my PhD program, and consequently this research, during the past few years.

A special thanks to my wife, Isabella and my family for supporting me during this "philosophical" period of my life.

10. References

[1] Bien, Z.Z., Lee, H.-E. (2007). "Effective learning system techniques for human-robot interaction in service environment." *Knowledge-Based Systems*, vol.20 no.5, pp.439-456.

[2] Billinghurst, M., Weghorst, S. (1995). "The use of sketch maps to measure cognitive maps of virtual environments." *Virtual Reality Annual International Symposium (VRAIS'95)*, pp.40.

[3] Bowman, D., Kruijff, E., LaViola Jr., J., Poupyrev, I. "3D User Interfaces: Theory and Practice" Addison-Wesley, Boston, 2005, ISBN 0-201-75867-9.

[4] Drury, J.L., Hestand, D., Yanco, H.A., Scholtz, J. (2004). "Design Guidelines for Improved Human-Robot Interaction." *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, pp. 1540-1540.

[5] Drury, J.L., Yanco, H.A., Howell, W., Minten, B., Casper, J. (2006). "Changing Shape: Improving Situation Awareness for a Polymorphic Robot." *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pp. 72-79.

[6] Johnson, C.A., Adams, J.A., Kawamura, K. (2003). "Evaluation of an Enhanced Human-Robot Interface." *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 900-905.

[7] Lindeman, R.W., Yanagida, Y., Noma, H., Hosaka, K. (2006). "Wearable Vibrotactile Systems for Virtual Contact and Information Display." *Special Issue on Haptic Interfaces and Applications, Virtual Reality*, 9 (2-3), pp. 203-213.

[8] Nielsen, C.W., Goodrich, M.A. (2006). "Comparing the Usefulness of Video and Map Information in Navigation Tasks." *Proceedings of the ACM Conference on Human-Robot Interaction*, pp. 95-101.

[9] Nielsen, C.W., Goodrich, M.A., Ricks, B. (2007). "Ecological interfaces for improving mobile robot teleoperation." *IEEE Transactions on Robotics*, 23 (5), pp. 927-941.

[10] Scholtz, J. (2002). "Evaluation methods for human-system performance of intelligent systems." *Proceedings of the 2002 Performance Metrics for Intelligent Systems (PerMIS) Workshop*.

[11] Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., Goodrich, M. (2006). "Common Metrics for Human-Robot Interaction." *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pp. 33-40.

[12] Yanco, H.A., Drury, J.L., Scholtz, J. (2004). "Beyond Usability Evaluation: Analysis of Human-Robot Interaction at a Major Robotics Competition." *Human-Computer Interaction*, 19 (1&2), pp.117-149.

[13] Yanco, H.A., Baker, M., Casey, R., Keyes, B., Thoren, P., Drury, J.L., Few, D., Nielsen, C., Bruemmer, D. (2006). "Analysis of Human-Robot Interaction for Urban Search and Rescue." *Proceedings of the IEEE Symp. on Safety, Security and Rescue Robots*.

Appendix A

The script used to guide the experimenter during the experiment is presented below.

Experimenter Script

Hello. My name is AAAAAA.

Welcome to the HIVE lab. Please, have a seat here.

Please, read the consent form and, if you agree, sign it at the bottom of the second page.

    {
    Subject reads and signs the consent form
    }

We are going to start with a few demographic questions:

How old are you?

How often do you play video-games: daily, weekly, seldom or never?

How often do you use or work with robots: daily, weekly, seldom or never?

Now, please carefully read these instructions, and let me know if you have any questions.

    {
    Subject reads the instruction sheet
    }

Any questions?

    {
    Answer questions asked
    }

I am now going to show you how to control the robot.

Using the left-hand analog stick of the controller, you can control the robot direction, making it move forward or backward, or making it turn left or right.

Using the right-hand analog stick you can tilt and pan the robot camera, whose movement is reflected on the video panel in front of the robot.

In order to take a picture, you first zoom in by pressing the trigger button #2 on the right side of the controller. You can adjust the picture by moving either the camera or the robot, but be careful with collisions.

When you are satisfied with the picture and while still pressing the zoom button, you press the trigger button #2 on the left side of the controller. This will take a screen shot of the robot camera current view. After that, you can simply release both trigger buttons and move on with your task.

You will have some time now to practice robot controls in a training room. The task here is to take a picture of a single red sphere, just like the ones you will have to locate in the real task, and which is hidden somewhere in this room. After that, you can keep practicing with the robot controls. Let me put this belt on you. It may or may not be activated. Activation is defined randomly by the application. Feel free to ask me questions about how to proceed with the study during this training session.

When you feel ready to start the search task, let me know and I will activate it. You will not be allowed to ask questions once the experiment starts, so please do so now.

{

Start chronometer for measuring training session time.

Training task is run until user requests to move on to the real task. Instructions during training session:

"The robot is represented by this red box in the middle of the screen. The blue lines represent the surfaces of objects that are close to the robot and that are detected by its sensors. Remember, if you try to move the robot and the robot does not move, it is because it is colliding with some object in the environment."

- If subject interface contains graphical feedback ring: "The set of cylinders around the robot give feedback on potential collisions that may occur in certain directions. The more red a cylinder gets, the closer you are to colliding in its direction. The indicators are not completely precise, so be careful."
- If subject interface (also) has the TactaBelt activated: "The belt around your waist is (also) providing you with feedback on potential collisions in certain directions. The more it vibrates in a certain direction, the closer you are to colliding with an object in that direction. (Once again,) The indicators are not completely precise, so be careful."

(Wait for the subject to find the sphere and take a picture. If he/she takes too long searching for it, provide him/her with hints so that he/she finds it quicker.)

"Do you have any questions on how to operate the robot?"

Stop chronometer.

}

Now I am going to start the real task. The objective of the task is to find as many red spheres as you can in as little time as possible and colliding the robot with the environment as little as you can. Once you are done with your search, you should move out of the house by passing through the exit door, which is identified by an exit sign on top of it, much like the one we have here in the lab. So please try to pay attention to that. Once you pass through the door, task will be over. I will start the task now, ok?

{

Start chronometer for measuring task time.

Task is run, no questions allowed.

Stop chronometer when task is over.

}

Now, please fill in this questionnaire. You can browse this document here with the pictures you have taken to help you with the description of the location of the red spheres you found. Feel free to use either pen or pencil.

{

Start chronometer for measuring sketching time.

Subject fills-in post-questionnaire.

Stop chronometer when sketching is over.

}

Do you have any other questions about this study or the lab?

Since other colleagues from your class might come to participate in this study, please avoid discussing what you did with others in order to avoid bias in our results, ok? Thank you very much for your participation.

Appendix B

The instructions contained in the instruction sheet that was used to explain the experiment to subjects is presented below.

Instructions

This experiment aims to evaluating the effect of a tactile interface on robot teleoperation.

Task: You will maneuver a virtual robot through a virtual environment, search for red spheres, and then maneuver the robot through an exit. You will have to do this task as fast and effectively as possible.

You will be presented with a house-like virtual environment. The house will have objects spread around in a chaotic manner so as to reproduce a catastrophic situation. Among the objects there will be red spheres. You will have to locate them by navigating a robot through the debris. Please memorize the locations of the spheres so that you can report them afterwards by sketching a map of the space and the sphere locations on a sheet of paper. You will be able to take screen captures of the location of the spheres that you can use later during sketching.

The world will be seen by using the robot camera present in the virtual robot interface. The camera will display the simulated real world. Other information obtained from the simulated real world will be displayed to you through the robot interface. You will be asked to perform the search task once. A timer will count the amount of time spent during task. The task will be over once you exit the house through the exit door, which is going to be identified by an emergency exit symbol.

The interface of the program contains a virtual representation of the robot and a virtual representation of the robot camera that displays images from the simulated real world. Additionally, a ring may be around the robot. If it is present, it will blink in different directions according to whether the robot is moving towards a direction that will cause imminent collision. In addition, you are wearing a belt with eight tactors. They may provide you with feedback on imminent collision situations with the robot. If the tactors are active, the closer the robot is to colliding, the more they will vibrate in the approximate direction of the collision. Objects that are close to the robot and detected by the sensors will also be virtually displayed by blue lines. The teleoperation interface therefore provides you with collision proximity detection, robot orientation and position, robot-camera orientation and identification of nearby objects.

It is important to notice that if you are trying to move the robot and it does not move, it is because the robot is colliding with objects in the real world.

Please sit comfortably during the experiment, but pay attention to the search task. After reading this, the experimenter will present you with the controls for the robot and give you time to get accustomed to them in a training room. If you have questions about how to proceed in the experiment, please, ask during the training session.

After that, feel free to ask the experimenter to start the experiment whenever you are ready. Once the actual experiment starts, please hold any questions until you are finished. Further information about the project will be given by the experimenter after you have finished the experiment. This is done in order not to bias the experiment results.

Appendix C

The post-questionnaire that was passed after subjects completed the task is presented below with the necessary blank space for answers removed.

Post-questionnaire

Subject #: ___

Please answer the questions in the empty space following them:

1. How many red spheres did you find?
2. In the space below, and using the pictures taken as a reference, please:
   a. Sketch a map of the space as you remember it, and
   b. Indicate on the sketch the location of each of the spheres.
3. Please provide any comments about the robot interface.
4. Do you have any comments about the experiment in general?
5. If you wish to know about the final results of this experiment, please, provide us with your e-mail address: ____.